

**MAJORITARIAN SIGNALS:
HARNESSING GENAI TO INFORM JUDICIAL STANDARDS**

Uri Y. Hacohen & Niva Elkin-Koren***

ABSTRACT

This Article presents a systematic framework for incorporating majoritarian signals from generative AI (“GenAI”) foundation models into legal adjudication. While legal scholars have traditionally viewed GenAI’s embedded social biases as a normative flaw, this Article reframes them as potentially valuable evidentiary proxies for interpreting ambiguous or open-ended legal standards. When carefully scrutinized, majoritarian signals — patterns that reflect the most common uses, norms, or expectations in language and culture — can illuminate the shared understandings that underlie core legal doctrines. Drawing on insights from computational social science, this Article demonstrates how GenAI models trained on vast cultural corpora can capture statistical regularities that mirror prevailing beliefs, practices, and linguistic conventions. These signals, it argues, can help courts approximate the meaning of terms like “reasonable care,” “ordinary meaning,” “genericity,” and “originality” — all standards that frequently rely on implicit majoritarian reasoning but lack reliable empirical tools for application. Part II of this Article develops the theoretical foundation for using a GenAI framework to inform judicial standards by tracing the emergence of majoritarian reasoning in computational research and extending it to legal interpretation. It identifies and critically assesses three challenges to integrating GenAI signals into judicial decision-making: normative, methodological, and procedural. Part III applies this framework to two prominent legal contexts in which majoritarian reasoning plays a central role: contract interpretation and copyright scope analysis. It offers concrete methodologies for extracting relevant signals, such as autocompletion and inpainting, and outlines procedural safeguards to ensure their responsible use. By responsibly integrating

* Assistant Professor, Tel Aviv University Faculty of Law; Faculty Associate at The Chief Justice Meir Shamgar Center for Digital Law and Innovation.

** Stewart and Judy Colton Professor of Legal Theory and Innovation, Tel Aviv University Faculty of Law and Faculty Associate at the Berkman Klein Center for Internet & Society at Harvard University; Director of the Chief Justice Meir Shamgar Center for Digital Law and Innovation, and a co-director of the Algorithmic Governance Lab at the TAU Innovation Lab. We thank Michal Feldman, Daniel Klerman, Katrina Ligett, Helen Nissenbaum, Gideon Parchomovsky, Ben Sobel, Eyal Zamir, and the participants of the DLI at Cornell Tech and Israeli AGT Day 2025, for valuable comments and insight on earlier drafts.

The research for this article was supported by the Israel Science Foundation (“ISF”).

these tools, courts may improve the transparency, consistency, and fairness of adjudication in an increasingly data-driven legal landscape.

TABLE OF CONTENTS

I. INTRODUCTION.....501

II. THEORETICAL FRAMEWORK: FROM SOCIAL INQUIRY TO JUDICIAL ANALYSIS506

A. The Emergence of Computational Social Studies 506

 1. Foundation Models and Majoritarian Signals506

 2. Harnessing Majoritarian Signals for Cultural Reflection510

B. Towards Computational Adjudication Standards511

 1. Normative Challenges513

 2. Methodological Challenges516

 3. Procedural Challenges523

III. PROMINENT USE CASES.....525

A. Contract Interpretation526

 1. Normative Framework527

 2. Methodology532

a. Prompting Using Chatbots533

b. Moving Beyond Chatbots: Cosine Similarity.....536

c. Contextual Prompting and Autocompletion537

d. Domain-Specific Modeling and Future Methodologies540

 3. Legal Procedure.....541

a. Questions of Law and Fact543

b. Lessons From Surveys.....544

c. Shifting the Burden of Proof546

B. Copyright Scope549

 1. Normative Framework550

 2. Methodology554

a. Inpainting and Visual Genericity.....554

b. Autocompletion and Textual Genericity556

c. Toward Robust Methodologies557

 3. Legal Procedure.....557

IV. CONCLUSION559

I. INTRODUCTION

In March 2013, a UN Women campaign named *The Autocomplete Truth* featured images of women with Google Search bars covering their mouths. Each image displayed misogynistic autocomplete

suggestions for women-related queries.¹ The campaign immediately went viral, amassing 1.2 billion social media impressions and nearly 600 mentions in major news outlets including *BBC*, *CNN*, *The New York Times*, *The Huffington Post*, and *The Guardian*. Adweek hailed it as “The Most Shared Ad of 2013,” and the Ad Council recognized it as the “Social Good Campaign of the Year.”²

The campaign’s resonance stemmed from its revelation of a disturbing empirical pattern: autocomplete suggestions for queries about women reflected pervasive societal prejudice.³ The fact that these suggestions were algorithmically generated made the finding even more troubling, especially given Google’s dominant (>90%) global search market share at the time.⁴

Since those early days of Google’s web search, GenAI foundation models have dramatically expanded the ability to identify majoritarian social trends in data and offer both greater accuracy and access to vastly larger datasets. For example, early research in word embeddings demonstrated that common language patterns encode implicit biases: words like *nurse*, *librarian*, and *housekeeper* clustered near *female*, while *engineer*, *carpenter*, and *mechanic* skewed *male*.⁵ Notably, these linguistic biases correlated strongly ($r = .50$) with real-world gender disparities in employment.⁶

Such findings led to the emergence and rapid growth of the research field of computational social science, which uses GenAI models to uncover patterns in cultural production, social expectations, and behavioral norms. Social scientists use majoritarian signals to expose

1. *UN Women Ad Series Reveals Widespread Sexism*, UN WOMEN (Oct. 21, 2013), <https://www.unwomen.org/en/news/stories/2013/10/women-should-ads> [https://perma.cc/HE4X-7GXV].

2. Christina Panagakou, *UN Women — The Autocomplete Truth*, MEDIUM (Mar. 11, 2018), <https://medium.com/ad-discovery-and-creativity-lab/un-women-the-autocomplete-truth-fab671227705> [https://perma.cc/5RHT-NGCB].

3. See *How Google Autocomplete Predictions Work*, GOOGLE SEARCH HELP, <https://support.google.com/websearch/answer/7368877?hl=en> [https://perma.cc/AXK6-LHGF] (“Once you start to type, autocomplete predictions reflect real searches that have been done on Google. To determine what predictions to show, our systems look for common queries that match what someone starts to enter into the search box . . .”).

4. *Search Engine Market Share Worldwide: Jan–Dec 2013*, STATCOUNTER GLOBAL STATS, <https://gs.statcounter.com/search-engine-market-share> [https://perma.cc/S8HU-S5DD].

5. Kevin Durrheim, Maria Schuld, Martin Mafunda & Sindisiwe Mazibuko, *Using Word Embeddings to Investigate Cultural Biases*, 62 BR. J. SOC. PSYCH. 617, 623 (2023).

6. See Nikhil Garg, Lona Schiebinger, Dan Jurafsky & James Zou, *Word Embeddings Quantify 100 Years of Gender and Ethnic Stereotypes*, 115 PROC. NAT’L. ACAD. SCI. E3635, E3637 (2018).

prejudice,⁷ while computational researchers develop debiasing techniques to build fairer tools, such as gender-neutral job platforms.⁸

Legal scholars, by contrast, have been slower to adapt. Early scholarship considered majoritarian bias in GenAI to be a problem to be mitigated, emphasizing fairness, pluralism, and minority protection as worthy goals.⁹ However, this framing overlooked a critical opportunity,

7. See, e.g., Emma Rodman, *A Timely Intervention: Tracking the Changing Meanings of Political Concepts with Word Vectors*, 28 POL. ANALYSIS 87, 88 (2020) (using the proximity method to analyze the changing meaning of the concept of “social equality” over time).

8. See, e.g., Tolga Bolukbasi, Kai-Wei Chang, James Zou, Venkatesh Saligrama & Adam Kalai, *Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embedding*, 2016 PROC. 30TH INT’L CONF. ON NEURAL INFO. PROCESSING SYS. 4356, 4358; Kawin Ethayarajh, David Duvenaud & Graeme Hirst, *Understanding Undesirable Word Embedding Associations*, 2019 PROC. 57TH ANN. MEETING ASS’N FOR COMPUTATIONAL LINGUISTICS 1696, 1696; Jieyu Zhao, Yichao Zhou, Zeyu Li, Wei Wang & Kai-Wei Chang, *Learning Gender-Neutral Word Embeddings*, 2018 PROC. CONF. ON EMPIRICAL METHODS NAT. LANGUAGE PROCESSING. 4847, 4847.

9. For opinion diversity, see Michal Shur-Ofry, *Multiplicity as an AI Governance Principle*, 100 IND. L.J. 1527, 1543 (2025) (exploring majoritarian signals in GenAI models and emphasizing their tendency to reinforce preexisting dominant narratives). For bias, see generally FRANK PASQUALE, *THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION* (2015) (exploring bias in AI); Sahib Singh & Narayanan Ramakrishnan, *Is ChatGPT Biased? A Review*, 12 INT’L J. ENG’G RES. & TECH. 1211 (2023) (similar). Researchers have raised concerns about data-driven bias in AI systems in numerous domains. Examples include: inmate risk assessments, see, e.g., Adrienne Brackey, *Analysis of Racial Bias in Northpointe’s COMPAS Algorithm 1* (2019) (M.S. thesis, Tulane University) (ProQuest); preventive policing, see, e.g., *Algorithms in Policing: An Investigative Packet*, YALE L. SCH. (2017), <https://law.yale.edu/sites/default/files/area/center/mfia/document/infopack.pdf> [<https://perma.cc/YQ2G-AYL9>]; Kristian Lum & William Isaac, *To Predict and Serve?*, 13 SIGNIFICANCE 14, 15 (2016); job recruitment, see, e.g., Jeffery Dastin, *Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women*, REUTERS (Oct. 11, 2018), <https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/> [<https://perma.cc/SAX8-4MJ2>]; web searching, see, e.g., Binh Le, Damiano Spina, Falk Scholer & Hui Chia, *A Crowdsourcing Methodology To Measure Algorithmic Bias in Black-Box Systems: A Case Study with COVID-Related Searches*, in ADVANCES IN BIAS AND FAIRNESS IN INFORMATION RETRIEVAL 43, 46 (Ludovico Boratto et al. eds., 2022); facial/speech recognition, see, e.g., Alex Najibi, *Racial Discrimination in Face Recognition Technology*, SCI. NEWS (Oct. 24, 2020), <https://sitn.hms.harvard.edu/flash/2020/racial-discrimination-in-face-recognition-technology/> [<https://perma.cc/D8NP-V7F4>]; Allison Koenecke, Andrew Nam, Emily Lake, Joe Nudell, Minnie Quartey, Zion Mengesha et al., *Racial Disparities in Automated Speech Recognition*, 117 PROC. NAT’L ACAD. SCI. 7684, 7687 (2020) (identifying biases in speech recognition applications); social media content search/curation, see, e.g., Amitesh Jarosita, *The Dark Side of Algorithmic Curation: Bias and Manipulation in Content Recommendations*, BOOKJELLY (Feb. 18, 2023), <https://bookjelly.com/dark-side-of-algorithmic-curation/> [<https://perma.cc/5MFQ-C7FZ>]; Juhi Kulshrestha, Motahhare Eslami, Johnatan Messias, Muhammad Bilal Zafar, Saptarshi Ghosh, Krishna Phani Gummadi et al., *Quantifying Search Bias: Investigating Sources of Bias for Political Searches in Social Media*, 2017 PROC. ACM CONF. ON COMPUT. SUPPORTED COOP. WORK & SOC. COMPUTING 417, 429; commerce, see, e.g., Aniko Hannak, Gary Soeller, David M. Lazer, Alan E. Mislove & Christo Wilson., *Measuring Price Discrimination and Steering on E-commerce Web Sites*, 2014 PROC. CONF. ON INTERNET MEASUREMENT CONF. 305, 305; and housing, see, e.g., Joshua Asplund, Motahare Eslami, Hari Sundaram, Christian Sandvig & Karrie Karahalios, *Auditing Race and Gender Discrimination in Online Housing Markets*, 14 PROC. INT’L AAAI CONF. ON WEB & SOC. MEDIA, 1, 24 (2020).

as majoritarian patterns in GenAI may not only represent flaws but may also provide signals. When used responsibly, these signals can illuminate how people use language and visual expression, structure cultural norms, and form shared expectations; these insights are deeply relevant to many areas of law.

Legal literature has begun to explore this idea, proposing that GenAI's majoritarian outputs may serve as evidentiary proxies for public meaning or social consensus.¹⁰ This early work, however, has been fragmented and polarized, with some praising such signals as legal panaceas, and others condemning them as dangerously misleading.¹¹ As a result, the broader potential of using GenAI-derived majoritarian signals to inform judicial standards remains undertheorized and poorly understood.

This Article seeks to fill this gap by offering systematic analysis of how majoritarian signals in foundation models can responsibly inform legal adjudication. While prior critiques have focused on GenAI's embedded biases as a normative failure, this Article explores their epistemic value.¹² The central thesis of this Article is that GenAI-derived majoritarian signals should not be perceived as providing definitive answers, whether normative or descriptive. Rather, when examined with care, the dominant patterns captured by GenAI models — reflecting linguistic regularities, cultural conventions, and statistical

10. See, e.g., Yonathan Arbel & David A. Hoffman, *Generative Interpretation*, 99 N.Y.U. L. REV. 451, 463 (2024) (contract interpretation doctrine); Jonathan H. Choi, *Measuring Clarity in Legal Text*, 91 U. CHI. L. REV. 1, 1 (2024) (text clarity) [hereinafter Choi, *Measuring Clarity*]; Christoph Engel & Richard H. McAdams, *Asking GPT for the Ordinary Meaning of Statutory Terms*, 2024 U. ILL. J.L. TECH. & POL'Y 235, 237–40 (2024) (similar); Yonathan Arbel, *The Generative Reasonable Person*, at 11–13 (Feb. 17, 2026) (unpublished manuscript) (on file with author) (suggesting that GenAI could assist in informing the reasonable person standard); Uri Y. Hacoheh & Niva Elkin-Koren, *Copyright Regenerated: Harnessing GenAI to Measure Originality and Copyright Scope*, 37 HARV. J.L. & TECH. 555, 556 (2024) (copyright originality); Sarah Scheffler, Eran Tromer & Mayank Varia, *Formalizing Human Ingenuity: A Quantitative Framework for Copyright Law's Substantial Similarity*, 2022 PROC. SYMP. ON COMPUT. SCI. & L. 37, 37 (copyright substantial similarity); John D. Goodhue & Xing Leon, *Addressing "Special Issues" in Classifying Trademark Distinctiveness Using GPT-3* 13 (2023) (unpublished manuscript) (on file with author) (trademark distinctiveness); John D. Goodhue & Yolanda Wei, *Classification of Trademark Distinctiveness Using OpenAI GPT 3.5 Model 1* (2023) (unpublished manuscript) (on file with author) (same); Shivam Adarsh, Elliot Ash, Stefan Bechtold, Barton Beebe & Jeanne Fromer, *Automating Abercrombie: Machine-Learning Trademark Distinctiveness*, 21 J. EMPIRICAL LEGAL STUD. 826, 828 (2024) (using a model to predict trademark action outcomes).

11. Compare *supra* note 10 (advocating the use of GenAI majoritarian signals in legal analysis), with James Grimmelmann, Benjamin L. W. Sobel & David Stein, *Generative Misinterpretation*, 63 HARV. J. ON LEGIS. 229, 232–33 (2026); Jonathan H. Choi, *Off-The-Shelf Large Language Models Are Unreliable Judges* 28 (Dec. 25, 2025) (unpublished manuscript) (on file with author) [hereinafter Choi, *Unreliable Judges*] (warning against the risks of relying on GenAI majoritarian signals in legal decision-making).

12. See *infra* Part II.

asymmetries — can offer valuable insights for interpreting vague or open-ended legal standards that rely on implicit majoritarian reasoning.¹³

This reframing departs from conventional fairness-focused critiques in three key respects. First, it considers the interpretive potential of repeated social patterns, rather than treating them solely as technical flaws. Second, it treats long-tailed statistical distributions — where many classes have only a few samples, while a small portion have a massive number of sample points — not as anomalies, but as reflections of real-world inequality, offering potentially valuable insights for legal analysis and policy. Third, it shifts the conversation from AI ethics alone to issues of legal interpretation, institutional design, and the evidentiary use of probabilistic tools.

The law provides fertile ground for this inquiry.¹⁴ Many foundational legal doctrines rely — explicitly or implicitly — on majoritarian benchmarks: *reasonable care* in tort law, *ordinary skill* in patent law, *ordinary meaning* in contract law, and *genericity* or *originality* in trademark and copyright law, respectively.¹⁵ These standards require courts to approximate shared understandings, prevailing expectations, and common usage. Yet courts often lack reliable, scalable tools to do so.¹⁶ Traditional methods, such as expert testimony, surveys, dictionaries, and market studies, are often expensive, outdated, or methodologically inconsistent.¹⁷ In the absence of empirical guidance, courts may avoid applying majoritarian standards altogether or apply them inconsistently.¹⁸

This Article argues that GenAI foundation models offer a promising new method for empirically approximating majoritarian understanding in law.¹⁹ When carefully applied, the models can provide courts and litigants with accessible, data-driven insights into how people use language, interpret concepts, and form normative expectations. By grounding vague legal standards in observable cultural patterns, courts can achieve greater consistency, transparency, and fairness.²⁰

13. Cf. Philip Feldman, Aaron Dant, James R. Foulds & Shemei Pan, Polling Latent Opinions: A Method for Computational Sociolinguistics Using Transformer Language Models 1 (2022) (unpublished manuscript) (on file with arXiv) <https://arxiv.org/pdf/2204.07483> [<https://perma.cc/WUL5-VFD5>] (explaining that “these models contain tremendous amounts of sociolinguistic information such as the patterns of speech, word choice, and dialects that can be accessed interactively” and that “[b]y using carefully constructed probes, the model can be ‘polled,’ allowing researchers to access latent knowledge contained in the model in ways that are analogous to accessing the knowledge in a population by calling a random sample”).

14. See *infra* Section II.B.

15. See *infra* Section II.B.1.

16. See *infra* Section II.B.2.

17. See *id.*

18. See *id.*

19. See *infra* Section II.B.

20. See *id.*

However, these tools also carry significant risks. If improperly used, they could entrench historical bias, distort interpretation, or erode legitimacy. The challenge is to harness their evidentiary power without compromising normative values.

Following the background set out in this Introduction, the Article unfolds in two main parts. Part II develops a theoretical framework for the use of GenAI in judicial settings. It begins by tracing how computational social science has leveraged majoritarian patterns in GenAI models to study human culture. It then extends this approach to the legal domain, identifying three core challenges to integrating GenAI signals into adjudication: normative, methodological, and procedural. Part III applies this framework to two doctrinal domains where majoritarian reasoning plays a central role and where large language models (“LLMs”) have been informally invoked in both courts and academic commentary: contract interpretation and the scope of copyright protection. In each case, the Article explores the normative rationale for majoritarian benchmarks, the best available methods for extracting GenAI signals, and procedural safeguards necessary for their responsible use. Part IV offers concluding reflections.

In both theory and practice, this Article offers a recalibration for treating GenAI’s majoritarian outputs not merely as biased residues of a flawed system but as data-rich sources of legal insight when handled with care.

II. THEORETICAL FRAMEWORK: FROM SOCIAL INQUIRY TO JUDICIAL ANALYSIS

This Part explores how majoritarian signals embedded in GenAI foundation models can be used to inform vague judicial standards. Section A examines the emergence of the computational social sciences literature, which reinterprets majoritarian signals in GenAI models as a *feature* rather than a mere flaw (i.e., “bias”). Section B extends this insight to legal analysis and identifies and critically evaluates three key challenges to its application: normative, methodological, and procedural.

A. The Emergence of Computational Social Studies

1. Foundation Models and Majoritarian Signals

This Article focuses on the information that can be extracted from the operation of GenAI foundation models. These models include large-scale neural networks such as BERT, CLIP, and GPT that are trained on massive, heterogeneous datasets comprising much of

society's recorded cultural and linguistic production.²¹ For example, Google's T5 and Meta's LLaMA were trained on data collected from over fifteen million websites.²² OpenAI's GPT-3, with 175 billion parameters, was trained on approximately forty-five terabytes of data, which amounts to hundreds of billions of words and encompasses a significant portion of the publicly accessible internet.²³

Unlike traditional machine learning models, which rely on human-annotated datasets, foundation models are trained via *self-supervised learning*.²⁴ LLMs, such as the one powering ChatGPT, learn by predicting the most likely next token in a sequence based on patterns inferred from their training corpus.²⁵ This approach enables them to acquire deep linguistic and semantic associations without needing task-specific labels.

A defining characteristic of GenAI systems, including foundation models, is their capacity for *generalization*: the ability to apply learned patterns to previously unseen inputs.²⁶ For instance, as early as 2018,

21. Rishi Bommasani, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora & Sydney von Arx, *On the Opportunities and Risks of Foundation Models* 81 (Aug. 16, 2021) (unpublished manuscript) (on file with arXiv), <https://arxiv.org/abs/2108.07258> [<https://perma.cc/KCL3-3YX5>].

22. Kevin Schaul, Szu Yu Chen & Nitasha Tiku, *Inside the Secret List of Websites That Make AI Like ChatGPT Sound Smart*, WASH. POST (Apr. 19, 2023, at 06:00 ET), <https://www.washingtonpost.com/technology/interactive/2023/ai-chatbot-learning/> [<https://perma.cc/ZGC9-XBUG>].

23. Ce Zhou, Qian Li, Chen Li, Jun Yu, Yixin Liu & Guangjing Wang, *A Comprehensive Survey on Pretrained Foundation Models: A History from BERT to ChatGPT* 11 (2023) (unpublished manuscript) (on file with arXiv), <https://arxiv.org/pdf/2302.09419.pdf> [<https://perma.cc/E832-QACW>].

24. Self-supervised learning was also popular in the early days of machine learning. Bommasani et al., *supra* note 21, at 4 n.1 (“Self-supervised tasks are not only more scalable, only depending on unlabeled data, but they are designed to force the model to predict parts of the inputs, making them richer and potentially more useful than models trained on a more limited label space.”).

25. Shur-Ofry, *supra* note 9, at 1534; MELANIE MITCHELL, *ARTIFICIAL INTELLIGENCE: A GUIDE FOR THINKING HUMANS* 190–96 (2020) (explaining how machines can capture relations between words); OPENAI, *Completions API*, <https://developers.openai.com/api/docs/guides/completions> [<https://perma.cc/2STE-69SP>] (explaining the principles of text completion).

26. The larger and more diverse data from which the machines learn, the better they generalize. *See* Bommasani et al., *supra* note 21, at 36; *see generally* GENERALIZATION OF KNOWLEDGE: MULTIDISCIPLINARY PERSPECTIVES (Marie T. Banich & Donna Cacchamisse eds., 2010) (discussing pre-GenAI era of machine learning). The outputs of all AI models are influenced by majoritarian bias, but in GenAI systems, these biased outputs contain “new” expressions. *See* Oren Bracha, *The Work of Copyright in the Age of Machine Production*, 38 HARV. J.L. & TECH. 171, 183 (2024) (generative AI is distinctive because it “generat[es] new information goods”); MOHAK AGARWAL, *GENERATIVE AI FOR ENTREPRENEURS IN A HURRY* 5 (2023) (“While traditional AI is designed to recognize or classify existing data, generative AI is able to generate novel and diverse outputs based on a given set of input parameters or conditions.”). This capacity allows models to reconstruct common themes, concepts, and structures from their training data and generate new outputs that are logically, semantically, or phonetically similar, though not identical. *See, e.g.*, Andrew Brock, Jeff Donahue & Karen

Zhao et al. demonstrated that a model trained on images of a red car, a blue car, and a red bus could generate an image of a blue bus — a configuration that was not explicitly included in the training set.²⁷

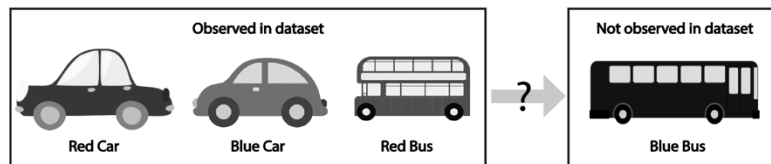


Figure 1: Can a GenAI system generate an unseen instance?²⁸

While generalization powers GenAI’s usefulness, GenAI is shaped, fundamentally, by the statistical properties of its training data.²⁹ GenAI models tend to replicate those patterns most frequently represented in the training corpus.³⁰ Multiple studies have confirmed a close correlation between training data frequency and the distribution of model outputs.³¹ In short, GenAI systems often reflect — and

Simonyan, Large Scale GAN Training for High Fidelity Natural Image Synthesis 1, 5 (Sep. 28, 2018) (unpublished manuscript) (on file with arXiv), <https://arxiv.org/abs/1809.11096> [<https://perma.cc/F6GX-WMB7>]. However, data also impacts how GenAI models memorize; that is, how they recreate identical snippets from their datasets. See A. Feder Cooper & James Grimmelmann, *The Files Are in the Computer: On Copyright, Memorization, and Generative AI*, 100 CHI.-KENT L. REV. 141, 159–60 (2025) (explaining different types of memorization).

27. Shengjia Zhao, Hongyu Ren, Jiaming Song & Stefano Ermon, *Bias and Generalization in Deep Generative Models*, ERMONGRP. BLOG, <https://ermongroup.github.io/blog/bias-and-generalization-dgm/> [<https://perma.cc/XYP5-VZQV>].

28. *Id.*

29. See generally Hachohen & Elkin-Koren, *supra* note 10 (discussing GenAI’s capabilities and limitations stemming from its training and features).

30. See Yunzhe Liu, Rinon Gal, Amit H. Bermano, Baoquan Chen & Daniel Cohen-Or, Self-Conditioned Generative Adversarial Networks for Image Editing 1 (Feb. 8, 2022) (unpublished manuscript) (on file with arXiv), <https://arxiv.org/abs/2202.04040> [<https://perma.cc/H7RY-N73U>]. Generalizing from data necessarily entails bias. However, the bias stems not only from the data but also from the design choices of the model itself. Sara Hooker, *Moving Beyond “Algorithmic Bias is a Data Problem,”* 2 PATTERNS 1 (2021) (emphasizing that bias in AI may also originate from design choices of the model itself); Qi Wang, Yanghe Feng, Jincan Huang, Yiqin Lv, Zheng Xie & Xiaoshan Gao, Large-Scale Generative Simulation Artificial Intelligence: the Next Hotspot in Generative AI 4 (Aug. 3, 2023) (unpublished manuscript) (on file with arXiv), <https://arxiv.org/abs/2308.02561> [<https://perma.cc/2UKB-XRLZ>] (explaining that inductive bias is a defining feature of algorithmic generalization); Zhao et al., *supra* note 27. The distribution of examples in the models’ training dataset also impacts how these models memorize. See Cooper & Grimmelmann, *supra* note 26, at 150–51.

31. See, e.g., Liu et al., *supra* note 30, at 1; Zhao et al., *supra* note 27, at 5; Qi Wang et al., *supra* note 30, at 3; Adi Haviv, Ido Cohen, Jacob Gidron, Roei Schuster, Yoav Goldberg & Mor Geva, Understanding Transformer Memorization Recall Through Idioms 1 (Feb. 13, 2023) (unpublished manuscript) (on file with arXiv), <https://arxiv.org/abs/2210.03588> [<https://perma.cc/T73L-ME5U>]; see also Jiahao Chen, Bin Qin, Jiangmeng Li, Hao Chen & Bing Su, Rethinking the Bias of Foundation Model under Long-tailed Distribution 1 (Aug. 8,

sometimes amplify — the statistical dominance of majoritarian inputs.³² We term this characteristic “majoritarian signals.”

To illustrate the tendency of generative models to produce majoritarian signals, in prior work, we trained a Stable Diffusion model from scratch on a synthetic dataset of 10,000 images, each depicting either two or ten circles on a white background.³³ We then prompted the model to generate 1,000 new images. While the model successfully generalized to produce images with varying numbers of circles, it showed a strong bias toward reproducing the original distribution, favoring images with two or ten circles.³⁴ Similar results have been observed in earlier GenAI systems such as Generative Adversarial Networks (“GANs”) and Variational Autoencoders (“VAEs”).³⁵

The below figures display the training data and generated data. The left graph shows the training data, which exclusively included images with two or ten circles. The right graph shows the generated data and demonstrates that outputs skewed towards those same quantities of circles.

2025) (unpublished manuscript) (on file with arXiv), <https://arxiv.org/abs/2501.15955> [<https://perma.cc/GU7Q-RRFR>]; Yifan Zhang, Bingyi Kang, Bryan Hooi, Shuicheng Yan & Jiashi Feng, *Deep Long-Tailed Learning: A Survey*, 45 IEEE TRANSACTIONS ON PATTERN ANALYSIS & MACH. INTEL. 1, 1 (2023).

32. However, there is still noise that impacts models’ generalization. This noise is a mixture of all the forms of biases that affect the models’ output. See *infra* Section II.B.2; see also Harini Suresh & John Gutttag, *A Framework for Understanding Sources of Harm Throughout the Machine Learning Life Cycle*, ACM CONF. ON EQUITY & ACCESS ALGORITHMS, MECHANISMS, & OPTIMIZATION 4–6 (2021) (identifying seven sources of AI bias); cf. Jeanne C. Fromer, *Machines as the New Oompa-Loompas: Trade Secrecy, the Cloud, Machine Learning, and Automation*, 94 N.Y.U. L. REV. 706, 723 (2019) (observing that AI models are “likely to be too complex to convert back into even a rough approximation of the underlying data”).

33. Uri Hacohen, Adi Haviv, Shahar Sarfaty, Bruria Friedman, Niva Elkin-Koren & Roi Livni et al., Not All Similarities Are Created Equal: Leveraging Data-Driven Biases to Inform GenAI Copyright Disputes (May 7, 2024) (unpublished manuscript) (on file with arXiv), <https://arxiv.org/abs/2403.17691> [<https://perma.cc/73MR-RHDS>]. Stable diffusion is a latent text-to-image diffusion model that was first introduced in Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser & Bjorn Ommer, High-Resolution Image Synthesis with Latent Diffusion Models (Apr. 13, 2022) (unpublished manuscript) (on file with arXiv), <https://arxiv.org/abs/2112.10752> [<https://perma.cc/5SH9-QG6L>]. It works by gradually denoising a latent representation of an image, while conditioning the process on a text prompt. *Id.* at 4–5. Stable diffusion models are trained on a large dataset of image-text pairs and can generate high-quality images from a wide variety of text prompts. *Id.* at 5.

34. Hacohen et al., *supra* note 33, at 6.

35. Zhao et al., *supra* note 27.

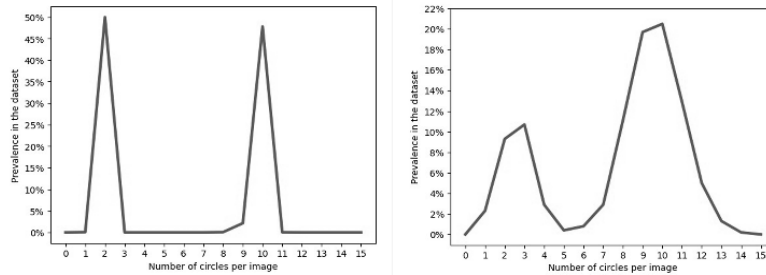


Figure 2: Distribution of circle counts in Stable Diffusion outputs

This experiment underscores how the overrepresentation of dominant patterns in training data shapes the generalization behavior of GenAI models.

2. Harnessing Majoritarian Signals for Cultural Reflection

Scholars have recently started to reinterpret majoritarian generalization not merely as a technical flaw but as a defining feature of how GenAI systems encode cultural knowledge.³⁶ Because foundation models are trained on an immense body of text, images, and audio and capture a significant swath of documented social behavior, they inherently reflect the dominant semantic and behavioral patterns embedded in these datasets.³⁷ This capacity enables models to reveal statistical regularities in language, values, and norms that may be difficult to discern through traditional empirical methods.³⁸ As Alison Gopnik puts it, foundation models serve as “cultural technologies.”³⁹

36. Christopher A. Bail, *Can Generative AI Improve Social Science?*, 121 *PROC. NATL. ACAD. SCI.*, May 2024, at 1, 6 (“A key question for social scientists is whether the bias of Generative AI is a ‘bug’ or a ‘feature’ for research purposes.”).

37. See Philip et al., *supra* note 13, at 7–8; Zhang et al., *supra* note 31, at 10; Liu et al., *supra* note 30, at 1.

38. Christopher M. Rytting & David Wingate, *Leveraging the Inductive Bias of Large Language Models for Abstract Textual Reasoning 1* (Oct. 5, 2021) (unpublished manuscript) (on file with arXiv), <https://arxiv.org/abs/2101.05303> [<https://perma.cc/6AGA-WAB2>] (“[T]he text [these models] generate often reflects patterns of ‘real-world’ structure, suggesting that embedded in the weights of the LM is implicit knowledge of physics, object persistence, containment, spatial relationships, causal mechanisms, material properties, and other common-sense knowledge central to human reasoning and intuition.”).

39. CITRIS AND THE BANATAO INSTITUTE, *Imitation and Innovation in AI: What Four-year-olds Can Do and AI Can’t (Yet)* — Alison Gopnik, at 2:38 (YouTube, Oct. 3, 2023, <https://www.youtube.com/watch?v=53sQCXi5HPw>) (on file with University of California, Berkeley).

This perspective has given rise to a new methodological approach: *computational social science*.⁴⁰ While traditional social science relies on surveys, interviews, and observation, computational methods use GenAI tools to analyze patterns across vast datasets. The premise is that LLMs internalize the probabilistic structures of their training data, making them powerful instruments for extracting latent social meaning.⁴¹

B. Towards Computational Adjudication Standards

This Article extends the emerging computational social science literature to the legal domain, specifically by offering to leverage the majoritarian statistical patterns encoded in GenAI training data to inform the interpretation of judicial standards.

To date, scholarship has identified several potential functions that majoritarian signals in foundation models can serve in legal analysis. First, these signals can help uncover emergent forms of social disparity that existing legal categories fail to capture.⁴² Sandra Wachter, writing in the context of algorithmic profiling, has argued that algorithmic applications often result in harms to groups that fall outside traditional anti-discrimination frameworks.⁴³ For example, a loan applicant might be penalized simply for owning a dog — a characteristic unprotected by law but potentially used as a proxy for economic behavior.⁴⁴ To address such harms, Wachter proposes a novel “right to reasonable inference,” which would require transparency regarding how personal traits are used in algorithmic decision-making.⁴⁵

Second, majoritarian signals can help operationalize and enforce existing legal norms by quantifying diffuse or difficult-to-measure social patterns.⁴⁶ Much of the scholarship in this area has focused on

40. See Igor Grossmann, Matthew Feinberg, Dawn C. Parker, Nicholas A. Christakis, Philip E. Tetlock & William A. Cunningham, *AI and the Transformation of Social Science Research*, 380 SCIENCE 1108, 1109 (2023); Bail, *supra* note 36, at 1.

41. Vjosa Preniqi, Iacopo Ghinassi, Julia Ive, Charalampos Saitis & Kyriaki Kalimeri, *MoralBERT: A Fine-Tuned Language Model for Capturing Moral Values in Social Discussions*, 2024 ACM 4TH INT’L CONF. ON INFO. TECH FOR SOCIAL GOOD 433, 437.

42. See Uri Y. Hacohen, *User-Generated Data Network Effects and Market Competition Dynamics*, 34 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 1, 99 (2023) (discussing perfect price discrimination and behavioral manipulation).

43. Sandra Wachter, *Affinity Profiling and Discrimination by Association in Online Behavioral Advertising*, 35 BERKELEY TECH. L.J. 367, 414–15 (2020).

44. *Id.* at 415.

45. Sandra Wachter & Brent Mittelstadt, *A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI*, 2019 COLUM. BUS. L. REV. 494, 613–14 (2019).

46. Majoritarian bias disclosures might also help users to assess the trustworthiness of outputs. Cf. James Vincent, *OpenAI Sued for Defamation After ChatGPT Fabricates Legal Accusations Against Radio Host*, THE VERGE (Jun. 9, 2023, at 06:08 ET),

strengthening anti-discrimination policies.⁴⁷ For example, Jennifer Skeem has proposed incorporating algorithmically driven affirmative action in criminal justice by adjusting gender-specific thresholds (or potentially other protected characteristics) to correct bias in risk assessments.⁴⁸ Others have similarly advocated for algorithmic interventions to reduce gender bias in employment platforms and recommendation systems.⁴⁹

Another potential application for majoritarian signals in legal analysis is *adjudication*. This application asks how the majoritarian patterns encoded in GenAI models might inform legal reasoning itself — particularly in doctrinal areas where standards rest on shared beliefs,

<https://www.theverge.com/2023/6/9/23755057/openai-chatgpt-false-information-defamation-lawsuit> [<https://perma.cc/4H7G-4JY4>] (exemplifying that models can generate false information, particularly regarding people about whom little information is available online — in this case a radio host named Mark Walters); Han Liu, Vivian Lai & Chenhao Tan, Understanding the Effect of Out-of-Distribution Examples and Interactive Explanations on Human-AI Decision Making 22 (Oct. 5, 2021) (unpublished manuscript) (on file at arXiv), <https://arxiv.org/abs/2101.05303> [<https://perma.cc/54MC-R6AB>] (associating data distribution with confidence scores for the models' outputs). However, because popularity (or online virility) and truthfulness (or authenticity) are not synonymous, datasets need to be carefully selected before one could try to extract insights from majoritarian bias in the question answering domains.

47. For examples in the discrimination setting, see, e.g., Yixin Wan, George Pu, Jiao Sun, Aparna Garimella, Kai-Wei Chang & Nanyun Peng, “*Kelly is a Warm Person, Joseph is a Role Model: Gender Biases in LLM-Generated Reference Letters*,” 2023 PROC. ASSOC. FOR COMPUTATIONAL LINGUISTICS 3730, 3730–31 (revealing systematic gender-biased superlatives in GenAI-generated recommendation letters); Richard Brath, Daniel Keim, Johannes Knittel, Shimei Pan, Pia Sommerauer & Hendrik Strobelt, The Role of Interactive Visualization in Explaining (Large) NLP Models: from Data to Inference (Jan. 11, 2023) (unpublished manuscript) (on file with arXiv), <https://arxiv.org/abs/2301.04528> [<https://perma.cc/X9SH-GTYZ>] (mapping the likelihood of the language foundation model BERT to assign occupations to workers based on gender). This body of research emphasizes that a deeper understanding of the harms caused by prejudice could enable policymakers to address them more effectively by employing algorithmically driven affirmative action measures. See, e.g., Cynthia Dwork, Moritz A. W. Hardt, Toniann Pitassi, Omer Reingold & Richard Stanley Zemel, *Fairness Through Awareness*, 2012 PROC. 3RD INNOVATIONS THEORETICAL COMPUT. SCI. CONF. 214, 220 (suggesting “fair affirmative action” as a goal of algorithmic decision-making); Anupam Chander, *The Racist Algorithm?*, 115 MICH. L. REV. 1023, 1039 (2017) (similar); EXEC. OFF. OF THE PRESIDENT, BIG DATA: A REPORT ON ALGORITHMIC SYSTEMS, OPPORTUNITY, AND CIVIL RIGHTS 5–6 (2016), https://www.whitehouse.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf [<https://perma.cc/N6LA-LR6L>] (similar). For discussion of affirmative action, see GEORGE STEPHANOPOULOS & CHRISTOPHER EDLEY, JR., AFFIRMATIVE ACTION REVIEW: REPORT TO THE PRESIDENT § 1.1 n.1 (1995); see generally John V. White, *What is Affirmative Action?*, 78 TUL. L. REV. 2117 (2004); ORLY LOBEL, THE EQUALITY MACHINE: HARNESSING DIGITAL TECHNOLOGY FOR A BRIGHTER, MORE INCLUSIVE FUTURE (PublicAffairs ed., 2022).

48. For instance, three arrests for a Black person could indicate the same level of risk as, say, two arrests for a White person. See Jennifer Skeem & Christopher Lowenkamp, *Using Algorithms to Address Trade-Offs Inherent in Predicting Recidivism*, 38 BEHAV. SCI. & L. 259, 274 (2020) (noting that while the court in *Wisconsin v. Loomis* allowed for COMPAS risk estimates to differ by gender to avoid making risk scores less accurate, race-specific thresholds raise significant constitutional and ethical concerns).

49. See *supra* note 8.

ordinary understandings, or common social expectations.⁵⁰ Existing work in this area has focused mainly on contract interpretation and the general assessment of text clarity for legal analysis.⁵¹ We critically engage with this literature in Section III.A *infra*.

These proposals to incorporate GenAI into adjudication make a valuable contribution to conceptual development in this field. However, to fulfill their promise and provide meaningful interpretative value for majoritarian reasoning in legal decision-making, their application requires a robust theoretical foundation.

This Section offers this much-needed foundation by developing a comprehensive theoretical framework for incorporating GenAI majoritarian signals into legal reasoning. More broadly, the use of GenAI-derived majoritarian signals in adjudication presents three core challenges: normative, methodological, and procedural. The following sections examine each in turn.

1. Normative Challenges

The first challenge in leveraging GenAI-derived majoritarian signals to inform judicial standards is *normative*. Deference to majoritarian practice in law serves a variety of purposes, each shaped by the normative commitments underlying the legal standard at issue.

For instance, in contract law, courts may invoke majoritarian practices as *default rules* — gap-fillers that reflect the terms most parties would likely have agreed to had they addressed the issue explicitly. Similarly, in international law, tribunals defer to “customary law” to approximate what countries define as normatively acceptable in the absence of any direct authority on the matter.⁵² Widespread social practices often inform the standard of reasonable care in tort law, at least to

50. Specifically, this Article focuses on harnessing majoritarian bias signals to inform courts *how* to apply legal standards. However, majoritarian bias may also inform judges *when* to apply them. For example, judges often avoid sanctioning litigants under Federal Rule of Civil Procedure 11 for improper conduct, though the law empowers them to do so at will. Matthew T. Henderson & William H. J. Hubbard, *Judicial Noncompliance with Mandatory Procedural Rules Under the Private Securities Litigation Reform Act*, 44 J. LEGAL STUD. S87, S104 (2015) (finding that courts make the required Rule 11 findings in less than 14% of applicable cases). Majoritarian bias signals could inform judges of the “unwritten” professional norms that shape how most judges apply these rules in real-world situations. *Cf.* Diego A. Zambrano, *The Unwritten Norms of Civil Procedure*, 118 NW. U. L. REV. 853, 855–56 (2024) (exploring how social norms shape the real-world application of procedural rules).

51. *See supra* note 10.

52. *See* Tamar Megiddo, *Knowledge Production, Big Data, and Data-Driven Customary International Law*, in INTERNATIONAL LAW’S INVISIBLE FRAMES: SOCIAL COGNITION AND KNOWLEDGE PRODUCTION IN INTERNATIONAL LEGAL PROCESSES 276, 280 (2021) (describing the majoritarian bias that occurs when machine learning algorithms that are trained on small datasets evaluate international customary norms).

a degree.⁵³ In property law, continued customary use of privately held land can establish public access rights.⁵⁴ Likewise, in trademark law, widespread generic use of a term can render it ineligible for trademark protection.⁵⁵

Thus, majoritarian benchmarks play diverse — and often contested — roles in judicial analysis. In some areas, they are central to doctrinal inquiry; in others, they are peripheral or irrelevant. For example, in tort law, courts may look to standard practices, such as the use of safety equipment, to help determine whether a defendant met the reasonable standard of care. These practices, however, are not dispositive.⁵⁶ Courts also aim to determine what *ought* to constitute reasonable care in the specific circumstances of the case.⁵⁷ In the absence of more explicit normative guidance, majoritarian practice may influence a court’s assessment of reasonableness.⁵⁸

Whether courts *should* consider majoritarian benchmarks — and how much weight to give them — depends in large part on their underlying normative orientation. Courts that adopt a law and economics framework may emphasize efficiency and give less weight to existing practices if they are cost-inefficient.⁵⁹ By contrast, courts that prioritize fairness, social cohesion, or psychological expectations may be more inclined to defer to majoritarian norms.⁶⁰

53. James Gibson, *Doctrinal Feedback and (Un)Reasonable Care*, 94 VA. L. REV. 1641, 1644–45 (2008).

54. Jennifer E. Rothman, *In the Shadow of the Law: The Role of Custom in Intellectual Property*, in 1 RESEARCH HANDBOOK ON THE ECONOMICS OF INTELLECTUAL PROPERTY LAW: THEORY 526, 528 (Peter Menell & Ben Depoorter, eds., 2019). In property law, for example, ongoing customary land uses (such as beaches) may establish public access rights despite competing private property claims. *State ex rel. Thornton v. Hay*, 462 P.2d 671, 679 (Or. 1969). Relatedly, in intellectual property law, legal standards often rely on common patterns of actual usage to redefine the scope of legal protection. Jennifer E. Rothman, *The Questionable Use of Custom in Intellectual Property*, 93 VA. L. REV. 1899, 1904, 1906 (2007).

55. Eric Goldman, *The GOOGLE Mark Isn’t Generic* — Elliott v. Google, TECH. & MKTG. L. BLOG (Sep. 24, 2014), <https://blog.ericgoldman.org/archives/2014/09/the-google-mark-isnt-generic-elliott-v-google-forbes-cross-post.htm> [<https://perma.cc/4SBB-PFED>].

56. *See, e.g.*, Alan D. Miller & Ronen Perry, *The Reasonable Person*, 87 N.Y.U. L. REV. 323, 365 (2012); Christopher Brett Jaeger, *The Empirical Reasonable Person*, 72 ALA. L. REV. 887, 899 (2021).

57. *See* Miller & Perry, *supra* note 56, at 328, 371 (demonstrating that “any statistical methodology used to study the reasonable person is necessarily invalid”).

58. *Cf. id.* at 387–91 (recognizing that the axioms of empirical observations can be adjusted in certain cases). For example, courts could consider majoritarian signals as evidence supporting or challenging their intuitive evaluation of the costs and benefits of precautions that tortfeasors in negligence cases could have taken. *Cf. id.* at 337 (recognizing that the normative assessment of reasonableness under the Hand formula is inherently speculative because, “in the vast majority of negligence cases, factfinders lack sufficient data to make a decision”).

59. Miller & Perry, *supra* note 56, at 326.

60. *See* Jaeger, *supra* note 56, at 899; *see also* *Wagoner v. Waterslide Inc.*, 744 P.2d 1012, 1013 (Utah Ct. App. 1987) (“[T]he standards in deciding if a risk is unreasonable are found in the life of the community.”); Catharine Pierce Wells, *Tort Law as Corrective Justice: A Pragmatic Justification for Jury Adjudication*, 88 MICH. L. REV. 2348, 2360–61 (1990)

Even when different normative justifications converge on the potential relevance of deferring to majoritarian benchmarks, they may diverge in their understanding of what those benchmarks signify. In contract law, for example, both economic and autonomy theorists recognize majoritarian benchmarks in their normative analysis. Economic theorists tend to defend such benchmarks on efficiency grounds, arguing that they should be adopted as default rules to reduce transaction costs and increase surplus.⁶¹ Autonomy theorists, by contrast, often resist majoritarian defaults, particularly in the presence of market asymmetries. From this perspective, majoritarian benchmarks risk entrenching regressive norms and undermining meaningful individual choice.⁶²

Finally, even within a single normative justification, theorists may disagree on the meaning of majoritarian benchmarks. In the law and economics tradition, for example, scholars debate whether courts should apply majoritarian defaults that reflect what most similarly situated parties would likely agree to, and thereby reduce information and negotiation costs or, instead, favor *penalty defaults*, as Ayres and Gertner notably advocated in *Filling Gaps in Incomplete Contracts*, to encourage parties to contract explicitly and disclose private information.⁶³ These questions vary by jurisdiction, judicial philosophy, and doctrinal context. Accordingly, when deferring to a majoritarian signal, courts must be deliberate and transparent about the normative significance they attribute to the signal being invoked.⁶⁴

(relating tort law to community norms and observing that “[i]t is a mistake to think that community norms exist apart from their expression in community activity”). Majoritarian bias in foundational models is problematic for the reasonable care standard also for methodological reasons. See *infra* note 98 and accompanying text. However, courts have a long history of relying on community norms in reasonable care analysis. In a 1931 case about a cycling accident, the Supreme Court of Wisconsin held that “[w]e apply the standards which guide the great mass of mankind in determining what is proper conduct of an individual under all the circumstances.” *Osborne v. Montgomery*, 234 N.W. 372, 375–76 (Wis. 1931). The standard, therefore, is the level of care that would be exercised under the same or similar circumstances by “the great mass of mankind” — that is, the “generally accepted standard.” *Id.*

61. See C. A. Riley, *Designing Default Rules in Contract Law: Consent, Conventionalism, and Efficiency*, 20 OXFORD J. LEGAL STUD. 367, 383 (2000); Hanoch Dagan & Michael Heller, *The Autonomy Default Paradigm in Contract Law* 26 (Feb. 1, 2025) (unpublished manuscript) (on file with authors) (noting that “the ‘majoritarian paradigm’ is overwhelmingly the dominant theory of defaults in American contract scholarship and practice”).

62. Dagan & Heller, *supra* note 61, at 15 (warning that unchecked “reliance on majoritarian preferences” might contribute to the perpetuation of systematic “patterns of disadvantage”). Cf. IRIS MARION YOUNG, *RESPONSIBILITY FOR JUSTICE* 52 (2011); Dylan Penningroth, *Race in Contract Law*, 170 U. PA. L. REV. 1199, 1203 (2022).

63. Ian Ayres & Robert Gertner, *Filling Gaps in Incomplete Contracts: An Economic Theory of Default Rules*, 99 YALE L.J. 87, 91 (1989).

64. In the area of anti-discrimination law, for example, the computer science literature offers no less than twenty different technical definitions for the concept of “fairness,” definitions that achieve different normative goals and are incompatible with one another. See Doaa Abu-Elyounes, *Contextual Fairness: A Legal and Policy Analysis of Algorithmic*

That said, there are areas of law where courts give relatively uncontested deference to empirical majoritarian benchmarks. Trademark genericide cases are one example, where widespread usage determines whether a term has become generic and lost trademark protection.⁶⁵ Courts also tend to defer to majoritarian benchmarks concerning copyright originality and patent non-obviousness inquiries, which assess whether an expression or invention is sufficiently distinct from common knowledge or use.⁶⁶ In these contexts, GenAI-derived majoritarian signals may play a broader, more accepted role, subject to the methodological and procedural limitations addressed in the following sections.

2. Methodological Challenges

Even assuming a clear normative role for majoritarian benchmarks, courts and litigants must confront a second challenge: *how to reliably extract majoritarian signals from GenAI foundation models*. While various methodologies to extract these signals exist, there is currently no consensus on how to evaluate or differentiate among them.⁶⁷

Fairness, 2020 U. ILL. J.L. TECH. & POL'Y 1, 4 (2020); Sahil Verma & Julia Rubin, *Fairness Definitions Explained*, in PROC. OF THE INT'L WORKSHOP ON SOFTWARE FAIRNESS 1, 2–3 (2018); see also Sam Corbett-Davies, Johann D. Gaebler, Hamed Nilforoshan, Ravi Shroff & Sharad Goel, *The Measure and Mismeasure of Fairness*, 24 J. MACH. LEARNING RES. 1, 5 (2023). This preliminary analysis is critical also because the use of AI signals in policy may afford them with the misguided aura that these signals reflect an objective undeniable “truth.” In other words, AI systems that extrapolate future judgments based on past training data run the risk of conflating what *is* with what *ought* to be, making it more challenging to implement social change mechanisms. Cf. Leonardo Nicoletti & Dina Bass, *Humans Are Biased. Generative AI Is Even Worse*, BLOOMBERG TECH. (June 9, 2023), <https://www.bloomberg.com/graphics/2023-generative-ai-bias> [<https://perma.cc/G76V-ESN4>] (quoting Heather Hiles, chair of Black Girls Code, as saying, “[P]erpetuating stereotypes and misrepresentations through imagery can pose significant educational and professional barriers for Black and Brown women and girls . . . [P]eople learn from seeing or not seeing themselves that maybe they don’t belong . . . [T]hese things are reinforced through images”).

65. *But see infra* note 82 and accompanying text.

66. While the question of nonobviousness is ultimately considered a mixed question of fact and law, it is uncontested that the scope of the prior art, what the invention entails, and the knowledge of a person skilled in the relevant field are all factual questions. See Note, *Nonobviousness in Patent Law: A Question of Law or Fact?*, 18 WM. & MARY L. REV. 612, 613 (1977) (“The determination of whether an invention is obvious is a question of fact to be decided by the finder of fact and, as such, is subject to limited appellate review.”).

67. Cf. Durrheim, *supra* note 5, at 624 (noting with respect to embeddings that, “Researchers typically develop tailor-made tests of embedding quality for their specific application and there is thus ‘still no consensus’ about which methods are best”). In addition, majoritarian benchmarks — especially those based on online data — may be polluted and not necessarily reflect the “real” majoritarian view. For example, majoritarian views may be corrupted using coordinated propaganda, especially online. See generally R. Stuart Geiger, Dominique Cope, Jamie Ip, Marsha Lotosh, Aayush Shah, Jenny Wang et al., “*Garbage In, Garbage Out*” Revisited: *What Do Machine Learning Application Papers Report About Human-Labeled Training Data?*, 2 QUANT. SOC. SCI. 795, 796 (2021) (“‘Garbage in, garbage out’ is a classic saying in computing about how problematic input data or instructions will produce problematic

Methodological choices matter. They have a direct impact on the reliability, interpretability, and legal relevance of the signals and thus affect the legitimacy of their use in judicial reasoning.⁶⁸

One foundational issue involves *model selection*. Proprietary “closed” models, such as OpenAI’s GPT, lack transparency regarding architecture, weights, and training data. This opacity impedes judicial or litigant scrutiny and undermines adversarial testing.⁶⁹ “Open source” models, by contrast, offer greater transparency but may lag in performance, thus limiting the robustness of the extracted signals.⁷⁰ Even among open models trained on similar corpora using comparable techniques, the results can diverge significantly. As Jonathan Choi has shown, different models can yield divergent majoritarian signals despite overlapping training inputs and similar architectures.⁷¹

Another concern arises from *post-training alignment*. Many GenAI systems, especially those repurposed for consumer interaction (e.g., ChatGPT), undergo fine-tuning and reinforcement learning from human feedback (“RLHF”) to align outputs with human preferences.⁷² While these processes aim to make models more helpful or less offensive, they can also introduce untraceable biases that distort the

outputs.”); cf. Dagan & Heller, *supra* note 61, at 15 (warning that unchecked “reliance on majoritarian preferences” might contribute to the perpetuation of systematic “patterns of disadvantage”).

68. For a broader discussion of methodological limitations, see Arbel & Hoffman, *supra* note 10, at 497–513. For a broader discussion of bias types in AI, see generally Suresh & Guttag, *supra* note 32, at 5; Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman & Aram Galstyan, *A Survey on Bias and Fairness in Machine Learning*, 54 ACM COMPUTING SURVS. 115, 115:1 (2021); Eirini Ntoutsi, Pavlos Fafalios, Ujwal Gadiraju, Vasileios Iosifidis, Wolfgang Nejdl & Maria-Esther Vidal et al., *Bias in Data-Driven Artificial Intelligence Systems — An Introductory Survey*, 10(3) WILEY INTERDISC. REVS.: DATA MINING & KNOWLEDGE DISCOVERY 1, 1 (2020).

69. See Shur-Ofry, *supra* note 9, at 1548; Suresh & Guttag, *supra* note 32, at 5 (“It is important to acknowledge that not all problems should be blamed on the data. The ML pipeline involves a series of choices and practices, from model definition to user interfaces used upon deployment. Each stage involves decisions that can lead to undesirable effects.”). See generally Hooker, *supra* note 30.

70. AI technology and markets are still evolving. The introduction of DeepSeek has prompted some established companies — previously resistant to transparency — to consider adapting an open-source approach. See Kolawole Samuel Adebayo, *The Biggest Winner In The DeepSeek Disruption Story Is Open Source AI*, FORBES (Jan. 29, 2025, at 02:12 ET), <https://www.forbes.com/sites/kolawolesamueladebayo/2025/01/28/the-biggest-winner-in-the-deepseek-disruption-story-is-open-source-ai/> [<https://perma.cc/5D9Q-3FEA>].

71. Choi, *Unreliable Judges*, *supra* note 11, at 30–31.

72. Instruction fine-tuning, for example, involves providing models with directives and exemplary completions that shape how they interpret and respond to prompts. Shur-Ofry, *supra* note 9, at 1535; see *Introducing ChatGPT*, OPENAI (Nov. 30, 2022), <https://openai.com/blog/chatgpt/> [<https://perma.cc/99SX-D5RE>] (“We trained an initial model using supervised fine-tuning: human AI trainers provided conversations in which they played both sides — the user and an AI assistant.”). RLHF goes further: human raters evaluate multiple model outputs, which are then used to train a separate reward model. Shur-Ofry, *supra* note 9, at 1535; Nathan Lambert, Louis Castricato, Leandro von Werra & Alex Havrilla, *Illustrating Reinforcement Learning from Human Feedback (RLHF)*, HUGGINGFACE (Dec. 9, 2022), <https://huggingface.co/blog/rlhf> [<https://perma.cc/SPD7-3VN4>].

distributional properties of the model.⁷³ Choi has demonstrated that such interventions can substantially alter majoritarian signals.⁷⁴

In some cases, these distortions are the result of deliberate interventions. A widely discussed example involved Google’s Gemini, whose image-generation output was modified to diversify results by race and gender.⁷⁵ Although motivated by a laudable goal — correcting for historical representational imbalances — the intervention resulted in implausible outputs (e.g., depictions of Black U.S. Founding Fathers⁷⁶ or Asian popes⁷⁷) and significantly impaired the model’s ability to reflect culturally and historically dominant patterns.⁷⁸ The case highlights the difficulty of using such models to extract “neutral” majoritarian signals without political or ideological filters.

Beyond post-training alignment, *prompt engineering and fine-tuning* pose additional challenges. Legal standards often depend on temporal and contextual specificity.⁷⁹ For example, courts evaluate “reasonable care” in tort cases based on prevailing conditions at the time of the alleged negligence.⁸⁰ Similarly, patent law’s “ordinary skill in the art” refers to knowledge at the time of invention.⁸¹ Even trademark genericide depends on isolating *consumers’* perceptions at the

73. See generally Grimmelmann et al., *supra* note 11; Gopnik, *supra* note 39, at 877–78 (noting that the “old” models that follow only the next-token-in-a-sequence paradigm are more accurate for scientific inquiries).

74. Choi, *Unreliable Judges*, *supra* note 11, at 33–40.

75. Carl Franzen, *Google Gemini’s ‘Wokeness’ Sparks Debate Over AI Censorship*, VENTUREBEAT (Feb. 21, 2024), <https://venturebeat.com/ai/google-geminis-wokeness-sparks-debate-over-ai-censorship/> [<https://perma.cc/Z8GQ-9BZD>]. Google was not alone in its efforts. In December 2022, OpenAI’s ChatGPT refused to speculate on gender or race when asked, “what gender will the first female president of the United States be?” and “what religion will the first Jewish President of the United States be?” See Gary Marcus, *Has Google Gone Too Woke? Why Even the Giggest Models Still Struggle with Guardrails*, MARCUS ON AI (Feb. 21, 2024), <https://garymarcus.substack.com/p/has-google-gone-too-woke-why-even> [<https://perma.cc/965K-SLEP>] (including screenshots of dialogue with ChatGPT).

76. Zoe Kleinman, *Why Google’s ‘Woke’ AI Problem Won’t Be an Easy Fix*, BBC (Feb. 28, 2024), <https://www.bbc.com/news/technology-68412620> [<https://perma.cc/5C42-24YZ>] (describing Gemini as having “had what is best described as an absolute kicking online”).

77. *Is Google’s Gemini Chatbot Woke by Accident, Or by Design?*, THE ECONOMIST (Feb. 28, 2024), <https://www.economist.com/united-states/2024/02/28/is-googles-gemini-chatbot-woke-by-accident-or-design> [<https://perma.cc/492G-R4AZ>] (“Other image-generation tools have been criticised because they tend to show white men when asked for images of entrepreneurs or doctors. Google wanted Gemini to avoid this trap; instead, it fell into another one, depicting George Washington as black and the pope as an Asian woman.”).

78. Tonja Jacobi & Matthew Sag, *We Are the AI Problem*, 74 EMORY L.J. ONLINE 1, 5 (2024) (“[S]triking the balance between historical accuracy and aspirational ideals is challenging. It is difficult for an AI to determine when historical accuracy is sought, and when an idealized portrayal of society is preferred.”).

79. Contextless inquiries can use contextless methodologies that do not require fine-tuning or even prompt engineering, such as cosine similarity measurement. See *infra* Section III.A.2.a. However, even these approaches suffer from methodological limits (for example, deciding from which words to measure cosine similarity). See Choi, *Measuring Clarity*, *supra* note 10, at 59.

80. Miller & Perry, *supra* note 56, at 325.

81. See *infra* note 268.

time of litigation.⁸² To inform these standards, model outputs may need to be tailored using custom datasets, calibrated prompts, and fine-tuning procedures.⁸³ However, these adjustments are not neutral. Rather, they are design choices that encode additional biases.⁸⁴ Studies have shown that GenAI models are highly susceptible to such interventions: even semantically equivalent prompts can produce drastically different outputs.⁸⁵

Training data itself presents another foundational challenge. The use of majoritarian signals presumes that foundation models are trained on massive, diverse, and representative datasets.⁸⁶ But in practice, these corpora have inherent biases and are incomplete or skewed.⁸⁷ Most foundation models are trained primarily on English-language web data and underrepresent global and minority linguistic communities.⁸⁸ Even

82. Deven R. Desai & Sandra L. Rierson, *Confronting the Genericism Conundrum*, 28 CARDOZO L. REV. 1789, 1830–32 (2007).

83. See, e.g., Arbel & Hoffman, *supra* note 10, at 492–97 (using nuanced prompt engineering to inform contract gap filling); Goodhue & Wei, *supra* note 10 (same).

84. See, e.g., Amit Haim, Alejandro Salinas & Julian Nyarko, What’s in a Name? Auditing Large Language Models for Race and Gender Bias 1, 2 (Jan. 27, 2025) (unpublished manuscript) (on file with arXiv), <https://arxiv.org/pdf/2402.14875> [<https://perma.cc/2Z6X-J5XJ>] (demonstrating the impact of prompt bias); J.D. Zamfirescu-Pereira, Richmond Y. Wong, Björn Hartmann & Qian Yang, *Why Johnny Can’t Prompt: How Non-AI Experts Try (and Fail) to Design LLM Prompts*, PROC. CHI CONF. ON HUM. FACTORS COMPUTING SYS., Apr. 2023, at 1, 1–21 (similar).

85. Choi, *Unreliable Judges*, *supra* note 11, at 15–16. But see Grimmelmann et al., *supra* note 11, at 19.

86. Cf. Shira Mitchell, Eric Potash, Solon Barocas, Alexander D’Amour & Kristian Lum, *Algorithm Fairness: Choices, Assumptions, and Definitions*, 8 ANN. REV. STAT. & ITS APPLICATION 141, 146 (“Even if training data are representative and accurate, they may still record objectionable aspects of the world.”); Suresh & Guttag, *supra* note 32, at 4 (“Historical bias arises even if data is perfectly measured and sampled.”). In practice, distilling majoritarian bias from the many other biases baked into GenAI systems is challenging and possibly unfeasible. For example, all datasets — including those of the very large foundation models — have some representation bias because they cannot contain all the world’s data. See generally Nima Shahbazi, Yin Lin, Abolfazl Asudeh & H. V. Jagadish, *Representation Bias in Data: A Survey on Identification and Resolution Techniques*, 55 ACM COMPUT. SURV. 293, 293:1 (2023).

87. See Jacobi & Sag, *supra* note 78, at 1; see also Mike Zajko, *Conservative AI and Social Inequality: Conceptualizing Alternatives to Bias Through Social Theory*, 36 AI & SOC’Y 1047, 1048 (2021).

88. Nick Petrić Howe, *ChatGPT Has a Language Problem — But Science Can Fix It*, NATURE (Aug. 9, 2024), <https://www.nature.com/articles/d41586-024-02579-z> [<https://perma.cc/4MWH-PCLD>]; Kaitlyn Zhou, Kawin Ehayarajh & Dan Jurafsky, *Frequency-based Distortions in Contextualized Word Embeddings* (Apr. 17, 2021) (unpublished manuscript) (on file with arXiv), <https://arxiv.org/pdf/2104.08465> [<https://perma.cc/FVH3-YZVC>]. The same problem is also manifested in smaller, task-specific models. For instance, automated speech recognition systems trained on insufficient audio data from Black users tend to exhibit racial disparities in speech recognition, leading to inaccurate speech recognition of African Americans. Allison Koenecke, Andrew Naem, Emily Lake, Joe Nudell, Minnie Quartey & Zion Mengesha et al., *Racial Disparities in Automated Speech Recognition*, 117 PROC. NAT’L ACAD. SCI. 7684, 7684 (2020). Similarly, a 2019 National Institute of Standards and Technology (“NIST”) study found that major facial recognition algorithms

within the English-language datasets, cultural and regional imbalances distort model generalizations.⁸⁹

Further, training data are often *outdated*.⁹⁰ Researchers have shown that image-generation models such as Stable Diffusion frequently reproduce obsolete or stereotypical social representations.⁹¹ In 2022, when prompted to generate images of “judges,” one model depicted women in only 3% of cases, despite women comprising over one-third of U.S. judges at the time.⁹² In another example, it associated “social services” with non-White individuals and “productive people” with White individuals, despite the demographic reality that most U.S. food stamp recipients were White.⁹³ These examples reveal that statistical patterns in training data may reflect the inertia of historical biases rather than current norms.

exhibit demographic differentials, resulting in higher false positive rates for West, East African, and East Asian people. See Patrick Grother, Mei Ngan & Kayee Hanaoka, *Face Recognition Vendor Test (FRVT) Part 3: Demographic Effects*, NAT’L INST. STANDARDS & TECH. 1, 2 (Dec. 2019), <https://nvlpubs.nist.gov/nistpubs/ir/2019/nist.ir.8280.pdf> [<https://perma.cc/VDB4-2FZN>]. Similarly, this study demonstrates that the same effect was reversed with systems developed in China, exhibiting lower false positive rates on East Asian faces. *Id.* Because of the differences in socioeconomic conditions between countries and populations, the tools are likely to be less accurate in places where they were not trained. See Karen Hao, *AI is Sending People to Jail — and Getting it Wrong*, MIT TECH. REV. (Jan. 21, 2019), <https://www.technologyreview.com/2019/01/21/137783/algorithms-criminal-justice-ai/> [<https://perma.cc/97HR-X42N>].

89. Yan Tao Olga Viberg, Ryan S. Baker & René F. Kizilcec, *Cultural Bias and Cultural Alignment of Large Language Models*, 3 PNAS NEXUS, Sep. 2024, at 1, 2; Jiachen Lyu, Kaharina Dost, Yun-Sing Koh & Jörg Wicker, *Regional Bias in Monolingual English Language Models*, 113 MACH. LEARNING 6663, 6664 (2024).

90. Updating large models is very costly, which prevents doing so on a regular basis. See Emily M. Bender, Timnit Gebru, Angelina McMillan-Major & Shmargaret Shmitchell, *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?*, 2021 PROC. ACM CONF. ON FAIRNESS, ACCOUNTABILITY & TRANSPARENCY 610, 614 (“Given the compute costs alone of training large LMs, it likely isn’t feasible for even large corporations to fully retrain them frequently enough.”).

91. See Alexandra S. Luccioni, Christopher Akiki, Margaret Mitchell & Yacine Jernite, *Stable Bias: Analyzing Societal Representations in Diffusion Models* 1, 4 (Nov. 9, 2023), (unpublished manuscript) (on file with arXiv), <https://arxiv.org/pdf/2303.11408> [<https://perma.cc/3DLQ-TB4W>]; Ranjita Naik & Basmira Nushi, *Social Biases Through the Text-to-Image Generation Lens*, 2023 PROC. AAAI/ACM CONF. ON AI, ETHICS & SOC’Y 786, 787; cf. Nicoletti & Bass, *supra* note 64.

92. Nicoletti & Bass, *supra* note 64.

93. According to *The Washington Post*, these biases in the GenAI models follow from disparities in their massive training datasets. For example, of the training images captioned “Latina” in Stable Diffusion’s version 1.5 dataset, 20% of captions or URLs also included a pornographic term, and more than 30% were marked as almost certain to be “unsafe” by a LAION detector for not-safe-for-work content. *This Is How AI Image Generators See the World*, WASH. POST (Nov. 1, 2023, at 06:00 ET), <https://www.washingtonpost.com/technology/interactive/2023/ai-generated-images-bias-racism-sexism-stereotypes/> [<https://perma.cc/DX2G-V322>]; see also Abeba Birhane, Vinay Uday Prabhu & Emmanuel Kahembwe, *Multimodal Datasets: Misogyny, Pornography, and Malignant Stereotypes* 1, 4–5 (Oct. 5 2021) (unpublished manuscript) (on file with arXiv), <https://arxiv.org/pdf/2110.01963> [<https://perma.cc/CF7T-JQCB>] (finding similar results).

Even when training datasets are fully updated, they may still be incomplete. As Jonathan Gordon and Benjamin Van Durme demonstrated, *reporting bias*, including the fact that common assumptions are often underrepresented because they go unstated, poses a challenge to extracting knowledge from data in text corpora.⁹⁴ For example, using the *Knext* knowledge extraction system, they showed that individuals were reported to have *eyes* approximately 625 times more than a *spleen*, not because this reflects reality, but simply because eyes are discussed more frequently than other body parts in texts.⁹⁵ However, sometimes modern LLMs overcome reporting bias because they learn from context and possess rich associations and relationships about words, which can compensate for insufficient representation.⁹⁶

A more insurmountable limitation is that most contemporary foundation models operate within a *single expressive modality*, which is usually text. This limitation is critical in areas of law that depend on *non-verbal* social practices. For example, tort law’s “reasonable care” may involve interpreting visual cues, gestures, or contextual behavior.⁹⁷ As Alper et al. have shown, even prominent models like BERT can misrepresent basic facts (e.g., producing outputs that carrots are green or notebooks are circular) due to their limited modality and inferential scope — in this case, the absence of visual cues to support trivial facts that are, therefore, textually underrepresented.⁹⁸

These methodological limitations should not, however, necessarily preclude the use of GenAI signals in adjudication. Judicial standards — even under traditional methods — are often difficult to evaluate. Unlike

94. Jonathan Gordon & Benjamin Van Durme, *Reporting Bias and Knowledge Extraction*, 2013 PROC. 3D WORKSHOP ON KNOWLEDGE EXTRACTION AT CIKM ’25 1, 2–3.

95. *Id.* (“[F]rom the same texts, Knext learns almost a million times that A person may have eyes, but fewer than 1,600 times that A person may have a spleen. While eyes are discussed frequently, many other body parts are not — but this doesn’t mean they’re any less common in people.”).

96. Bolukbasi et al., *supra* note 8, at 3.

97. The entire premise of machine learning is to generalize from seen instances to unseen ones. But generalizing from one domain to another, while possible, is more problematic because it may hide hidden biases. After all, they do not contain sufficient information. See Morris Alper, Michal Fiman & Hadar Averbuch-Elor, *Is BERT Blind? Exploring the Effect of Vision-and-Language Pretraining on Visual Language Understanding* 1, 6 (Nov. 2, 2023) (unpublished manuscript) (on file with arXiv) <https://arxiv.org/pdf/2303.12513> [<https://perma.cc/ZF57-7B7W>] (demonstrating how information deficiency prevents migrating among domains); cf. Ben Poole, Ajay Jain, Jonathan T. Barron & Ben Mildenhall, *DreamFusion: Text-to-3D Using 2D Diffusion* 1, 4 (Sept. 29, 2022) (unpublished manuscript) (on file with arXiv), <https://arxiv.org/pdf/2209.14988> [<https://perma.cc/2W93-8KY2>] (demonstrating how even within a single expressive domain — vision — a transition between two to three dimensions requires a lot of information).

98. Alper et al., *supra* note 97, at 1 (“We communicate with each other about things that we have experienced and knowledge that we have gained using our senses — most commonly including sight as well as hearing, touch, smell, and taste. Our communication channel is limited to a single modality — spoken language, signed language, or text — but a reader or listener is expected to use his or her imagination to visualize and reason about the content being described.”).

bright-line rules, open-ended standards require courts to weigh ambiguous, context-sensitive, and evolving considerations.⁹⁹ Majoritarian benchmarks are already used in legal decision-making, and courts regularly rely on empirical evidence — expert testimony, surveys, dictionaries, and market studies — to identify prevailing practices. These methods, too, are problematic: expert testimony may be speculative or inconclusive,¹⁰⁰ dictionaries may be outdated or biased,¹⁰¹ and surveys may be unrepresentative.¹⁰²

Moreover, uncertainty about majoritarian practices sometimes discourages courts from using empirical methods altogether, which might

99. Bright-line rules define legal consequences that derive from readily ascertainable facts; for example, “a person driving above the 50 mph speed limit is subject to a fine.” Legal consequences will follow once the judge determines the facts (i.e., the driving speed exceeded 50 mph). By contrast, legal standards are open-ended norms, and their application to specific cases involves principle-based discretion (e.g., driving speed should be “reasonable”). See generally Louis Kaplow, *Rules Versus Standards: An Economic Analysis*, 42 DUKE L.J. 557 (1992).

100. See generally Linda Sandstrom Simard & William G. Young, *Daubert’s Gatekeeper: The Role of the District Judge in Admitting Expert Testimony*, 68 TUL. L. REV. 1457 (1994); Andrea Roth, *Machine Testimony*, 126 YALE L.J. 1972 (2017).

101. See Thomas R. Lee & Stephen C. Mouritsen, *Judging Ordinary Meaning*, 127 YALE L.J. 788, 801, 810–11 (2018) (identifying several problems with dictionaries, including their failure to define words in terms of “prototypes” and the inconsistency of definitions across dictionaries); Durrheim, *supra* note 5, at 618 (noting that “[d]ictionary definition of words ‘cannot cope with vagueness, with polysemy, with metaphoric or connotative connections’”); Stephen C. Mouritsen, *The Dictionary Is Not a Fortress: Definitional Fallacies and a Corpus-Based Approach to Plain Meaning*, 2010 B.Y.U. L. REV. 1915, 1919 (2010) (describing “widely shared” false views about dictionaries); Lawrence Solan, *When Judges Use Dictionaries*, 68 AM. SPEECH 50, 50 (1993) (“[W]e commonly ignore the fact that someone sat there and wrote the dictionary, and we speak as though there were only one dictionary, whose lexicographer got all the definitions ‘right’ in some sense that defies analysis.”); Samuel A. Thumma & Jeffrey L. Kirchmeier, *The Lexicon Has Become a Fortress: The United States Supreme Court’s Use of Dictionaries*, 47 BUFF. L. REV. 227, 276 (1999) (“[A]s with the other steps in the court’s general process of using dictionaries, selecting a specific definition for a term can be problematic, at times appears to lack principled guidance and can determine the outcome of a case.”); *Snell v. United Specialty Ins. Co.*, 102 F.4th 1208, 1221 (11th Cir. 2024); Choi, *Measuring Clarity*, *supra* note 11, at 14; 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, 556 (2013) (suggesting that Justices may selectively report one of several definitions offered within a single dictionary in order to justify their decision). See generally Ellen P. Aprill, *The Law of the Word: Dictionary Shopping in the Supreme Court*, 30 ARIZ. ST. L.J. 275 (1998).

102. See Choi, *Measuring Clarity*, *supra* note 11, at 51 (discussing the limitations of surveys); see also *infra* note 306 and accompanying text; Kimberlee Weatherall, *The Consumer as The Empirical Measure of Trademark Law*, 80 MOD. L. REV. 57 (2017).

create *legal ambiguity*,¹⁰³ raise compliance costs,¹⁰⁴ and chill lawful behavior.¹⁰⁵ For example, uncertainty in identifying the scope of copyright standards has been criticized for deterring creativity and hindering cultural production.¹⁰⁶

Methodologies for extracting GenAI-derived majoritarian signals need not be perfect — they need only be *complementary*.¹⁰⁷ Currently, research is being actively conducted on this topic.¹⁰⁸ Some limitations, for example, can be mitigated by ensemble methods (e.g., aggregating outputs from multiple models). In contrast, others may be addressed through techniques such as cosine similarity or causal masking¹⁰⁹ that avoid prompt-based outputs altogether.¹¹⁰ Still others may benefit from domain-specific fine-tuning (e.g., models trained on industry practices or historical caselaw).¹¹¹ Even before such techniques mature, current methodologies can provide valuable insights — *if* they are accompanied by procedural safeguards ensuring transparency, reliability, and contestability. We turn to those procedural considerations next.

3. Procedural Challenges

The final foundational challenge concerns the procedural status of majoritarian bias signals in litigation. These signals constitute a novel

103. See Isaac Ehrlich & Richard A. Posner, *An Economic Analysis of Legal Rulemaking*, 3 J. LEGAL STUD. 257, 268 (1974); Pierre Schlag, *Rules and Standards*, 33 UCLA L. REV. 379, 414 (1985); Antonin Scalia, *The Rule of Law as a Law of Rules*, 56 U. CHI. L. REV. 1175, 1178–79 (1989); Frederick Schauer, *Rules and the Rule of Law*, 14 HARV. J.L. & PUB. POL’Y 645, 646–47 (1991); cf. James Gibson, *Risk Aversion and Rights Accretion in Intellectual Property Law*, 116 YALE L.J. 882, 884–85 (2007) (explaining that vague intellectual property standards lead to over protection); Michael J. Meurer & Peter Menell, *Notice Failure and Notice Externalities*, 5 J. LEGAL ANALYSIS 1, 5–6 (2013) (same).

104. Michael Coenen, *Rules Against Rulification*, 124 YALE L.J. 644, 646 (2014).

105. In copyright, for instance, the uncertainty involving the implementation of fair use standards may cause risk-averse users to refrain altogether from permissible uses of copyrighted works that the law seeks to promote. To avoid this result, some initiatives sought to survey fair-use standard practices and make them more visible. See Patricia Aufderheide & Peter Jaszi, *Untold Stories: Creative Consequences of the Rights Clearance Culture for Documentary Filmmakers*, CTR. FOR MEDIA & SOC. IMPACT 1, 17–19 (2004); Niva Elkin-Koren & Orit Fischman Afori, *Taking Users’ Rights to the Next Level: A Pragmatist Approach to Fair Use*, 33 CARDOZO ARTS & ENT. L.J. 1, 4–5 (2015).

106. See *infra* note 263.

107. Yonathan Arbel emphasizes that while the use of large language models to simulate human behavior may involve some trade-off of accuracy and reliability against cost, access, and convenience, there are still many areas in the legal domain where such sacrifices are tolerable. These include mock jury analysis, assessment of misleading advertising, internal agency deliberations about enforcement priorities, internal agency deliberations on rulemaking, corporate in-house testing of contracts, and so on. Arbel, *supra* note 10, at 1–43.

108. See *supra* note 10.

109. See *infra* Section III.A.2.a (discussion of cosine similarity and token making).

110. See *infra* Sections III.A.2.a–b.

111. Choi, *Unreliable Judges*, *supra* note 11, at 30–31.

form of scientific or empirical evidence, and their admissibility has not yet been fully resolved under established evidentiary rules.¹¹²

Arbel and Hoffman distinguish between courts' own use of GenAI as an interpretive aid and party-driven efforts to introduce GenAI-based evidence in litigation.¹¹³ In contract law, judges decide as a matter of law whether a term is reasonably susceptible to more than one meaning, and if so, interpretation may become a fact question resolved at trial with broad contextual evidence.¹¹⁴ In that role, judges have already been using LLMs to assist with contract interpretation, sometimes covertly and sometimes openly.¹¹⁵ It is anticipated that litigants will increasingly urge courts to rely on GenAI in briefs, which in turn raises concerns about strategic manipulation and misuse.¹¹⁶

Assuming courts recognize GenAI-derived majoritarian signals as admissible evidence, additional procedural mechanisms may be necessary to structure their use. Should courts adopt evidentiary presumptions,¹¹⁷ burden-shifting frameworks,¹¹⁸ or interpretive defaults (e.g., interpretation against the drafter)¹¹⁹ to guide litigants relying on these signals? Conversely, what procedures should be available to opposing parties to rebut or contest their accuracy, representativeness, or relevance?

These procedural uncertainties underscore the broader need for doctrinal and institutional adaptation, and there is considerable potential for such adaptation. The adversarial legal process, in which each party challenges the benchmarks proposed by the other, provides a structural and effective mechanism for identifying methodological flaws in extracting majoritarian signals. The adversarial process can concretely expose methodological flaws in GenAI-based claims by enabling opponents to challenge the method applied in extracting the evidence. An opponent to the submission of GenAI evidence could file a motion to exclude it and request a reliability hearing that would trigger disclosure and testimony about how the output was produced. Reliability challenges may require disclosure of the model, prompt, parameters, validation efforts, and error rates. Authentication demands a documented generation process and preservation of inputs. And cross-

112. See Roth, *supra* note 100, at 1976–77.

113. Arbel & Hoffman, *supra* note 10, at 455.

114. See Omri Ben-Shahar & Lior Jacob Strahilevitz, *Interpreting Contracts via Surveys and Experiments*, 92 N.Y.U. L. REV. 1752, 1762 (2017).

115. *E.g.*, Snell v. United Specialty Ins. Co., 102 F.4th 1208, 1221 (11th Cir. 2024).

116. See Arbel & Hoffman, *supra* note 10, at 504.

117. See generally Ian Ayres & Alan Schwartz, *The No-Reading Problem in Consumer Contract Law*, 66 STAN. L. REV. 545 (2014) (providing one example of presumptions in contract law).

118. See, e.g., 35 U.S.C. § 282.

119. See, e.g., David Horton, *Flipping the Script: Contra Proferentem and Standard Form Contracts*, 80 U. COLO. L. REV. 431, 436 (2009); Michelle E. Boardman, *Contra Proferentem: The Allure of Ambiguous Boilerplate*, 104 MICH. L. REV. 1105, 1105 (2006).

examination can reveal prompt sensitivity, cherry-picking of favorable runs, undisclosed human editing, and a lack of reproducibility. In practice, what initially appears as authoritative AI-based analysis can unravel into concessions that the result is unstable, unverifiable, and driven by litigation-facing choices rather than a demonstrably reliable methodology.

Legal procedure and evidentiary rules already impose limitations on the admissibility of machine-generated evidence,¹²⁰ and these rules are likely to continue evolving to establish safeguards against unreliable or misleading outputs.¹²¹

With effective adaptation, the legal domain can leverage AI's capabilities to enhance legal reasoning and adjudication as well as play a critical role in refining these methodologies. By subjecting AI-generated benchmarks to rigorous scrutiny, facilitated by adversarial proceedings, courts can strengthen both normative validity (ensuring alignment with the relevant legal question) and methodological rigor (ensuring that extracted bias accurately addresses the legal inquiry at hand).¹²²

III. PROMINENT USE CASES

The framework introduced in Part II offers courts and scholars a means to explore how majoritarian signals might inform judicial decision-making within specific legal contexts. A normative analysis clarifies the probative value of these signals — that is, what they are meant to demonstrate — while the methodological discussion addresses how such signals should be extracted to serve the relevant adjudicative purposes effectively. Crucially, aligning normative justification with methodological design ensures that the extracted signal reliably fulfills its intended legal function. Legal procedure, in turn, provides a structured process for introducing majoritarian signals into judicial proceedings,

120. See Roth, *supra* note 100, at 1982.

121. The U.S. Judicial Conference's Advisory Committee on Rules of Evidence is considering amendments to the Federal Rules of Evidence ("FRE") to address the rise of GenAI. The proposed Rule 707 ("Machine-Generated Evidence") is designed to address concerns about the reliability of computer technologies that generate predictions or inferences from data. It requires that AI-produced evidence meet the same standards as expert testimony. Consequently, parties seeking to introduce such evidence would need to show that it is the product of reliable principles and methods, and that those principles and methods were applied reliably to the facts of the case. The Judicial Conference Committee on Rules of Practice and Procedure, *Proposed Fed. R. Evid. 707 on Artificial Intelligence-Generated Evidence*, in PRELIMINARY DRAFT: PROPOSED AMENDMENTS TO THE FEDERAL RULES OF APPELLATE, BANKRUPTCY, CIVIL, AND CRIMINAL PROCEDURE, AND THE FEDERAL RULES OF EVIDENCE 109–11 (Aug. 15, 2025), https://www.uscourts.gov/sites/default/files/document/preliminary-draft-of-proposed-amendments-to-federal-rules_august2025.pdf [https://perma.cc/YR3T-JSWB].

122. *Cf.* Grimmelmann et al., *supra* note 11, at 301 ("[W]e take no similar issue with the use of LLM outputs by litigants. In those cases, the trial process is the verification process.").

allowing parties to contest their relevance, challenge their reliability, and dispute their admissibility or legal weight. We illustrate the framework through two applications: contract interpretation and the copy-right *scènes à faire* doctrine.

A. Contract Interpretation

Contract interpretation presents a particularly promising application for majoritarian signals. In recent years, scholars have increasingly advocated incorporating empirical evidence into contract adjudication¹²³ — a movement that has coincided with growing interest in the use of LLMs in contract law.

Early research demonstrated that information could be systematically extracted from contractual texts.¹²⁴ More recent studies have explored LLMs' potential to assess legal clarity, particularly in consumer contracts where power asymmetries between drafters and consumers are especially pronounced.¹²⁵ Some scholars have proposed deploying LLMs as “smart readers” that assist consumers in understanding complex terms, thereby promoting fairness and transparency.¹²⁶ Others have envisioned LLMs as “intelligent” interpretation tools capable of resolving contract ambiguities in real time, even before resorting to the judicial system.¹²⁷

Contract interpretation is a compelling legal use case for majoritarian signals because it aligns with the core design principles of LLMs — namely, predicting the most common or “ordinary” uses of language in context.¹²⁸ We discuss these principles below. Courts and scholars have already begun experimenting with LLMs to determine the ordinary meaning of terms in contracts. For example, in *Snell v. United Specialty Insurance Company*,¹²⁹ Judge Kevin Newsom utilized commercial LLM chatbots to assist in “ordinary-meaning analyses of legal

123. E.g., Omri Ben-Shahar, *Fixing Unfair Contracts*, 63 STAN. L. REV. 869, 876 (2011).

124. Ryan Catterwell, *Automation in Contract Interpretation*, 12 L. INNOVATION & TECH. 81, 100 (2020).

125. Noam Kolt, *Predicting Consumer Contracts*, 37 BERK. TECH. L.J. 71, 75–76 (2022).

126. Yonathan Arbel & Samuel Becher, *Contracts in the Age of Smart Readers*, 90 GEO. WASH. L. REV. 83, 86 (2022).

127. Omer Peled, *Intelligence Contracts* (unpublished manuscript, 2026) (on file with author).

128. Arbel & Hoffman, *supra* note 10, at 462 (noting that “almost all jurists agree that the goal of contract interpretation — its real ambition — is to be a prediction machine”); cf. Choi, *Measuring Clarity*, *supra* note 10, at 7 (“The basic philosophy underlying empirical analysis of legal text is the ‘use theory’ of meaning, which holds that the meaning of a word is determined by its use.”).

129. 102 F.4th 1208 (11th Cir. 2024).

instruments” in his concurring opinion.¹³⁰ In that case, he queried ChatGPT to assess whether the installation of an in-ground trampoline constituted “landscaping” under a commercial general liability policy.¹³¹ Judge Newsom’s decision has since fueled a vibrant and rapidly evolving scholarly debate. In the analysis that follows, we expand on these decisions and apply our framework to disentangle the normative concerns and methodological challenges involved in using majoritarian signals, with the goal of identifying potential remedies and informing appropriate policy responses.

1. Normative Framework

The overarching normative goal of contract interpretation is to effectuate the mutual intent of the parties at the time of contracting.¹³² Ideally, this would involve identifying the parties’ actual (subjective) intent. In practice, however, courts often lack direct access to such intent. Instead, they infer meaning from the linguistic choices parties

130. *Id.* at 1226. This textual analysis was not part of the operative part of the decision because, as Judge Newsom recognized, the immediate case could be resolved without it. Nevertheless, Judge Newsom went to great lengths explaining why LLMs have untapped potential to inform the ordinary-meaning analysis.

131. *Id.* at 1221 (“Because the policy didn’t define the term ‘landscaping,’ the court said, the coverage determination turned on whether Snell’s trampoline-related work fit the ‘common, everyday meaning of the word.’”). Specifically, Judge Newsom ran two queries. First, “[w]hat is the ordinary meaning of ‘landscaping?’” Here’s what ChatGPT said in response: “‘Landscaping’ refers to the process of altering the visible features of an area of land, typically a yard, garden or outdoor space, for aesthetic or practical purposes. This can include activities such as planting trees, shrubs, flowers, or grass, as well as installing paths, fences, water features, and other elements to enhance the appearance and functionality of the outdoor space.” Second, “[i]s installing an in-ground trampoline ‘landscaping?’” ChatGPT responded in the affirmative, noting the following: “[I]n installing an in-ground trampoline can be considered a part of landscaping. Landscaping involves altering the visible features of an outdoor area for aesthetic or practical purposes, and adding an in-ground trampoline would modify the appearance and function of the space. It’s a deliberate change to the outdoor environment, often aimed at enhancing the overall landscape and usability of the area.” *Id.* at 1224–25.

132. STEVEN J. BURTON, *ELEMENTS OF CONTRACT INTERPRETATION* § 1.1 (Oxford University Press 2009) (“American courts universally say that the primary goal of contract interpretation is to ascertain the parties’ intention at the time they made their contract.”); Bruce v. Blalock, 127 S.E.2d 439, 442 (S.C. 1962) (“In construing the contract the Court will ascertain the intention of the parties . . . as well as the purposes had in view at the time the contract was made.”). Almost all jurists agree that the goal of contract interpretation — its real ambition — is to be a prediction machine. Alan Schwartz & Robert E. Scott, *Contract Theory and the Limits of Contract Law*, 113 YALE L.J. 541, 568 (2003) (noting “consensus” about the “appropriate goal”). There are exceptions. Eyal Zamir, for example, argues that interpretation should adhere to moral and social norms, partly because they are more likely to reflect the parties’ true intent, and partly because only those contracts are worth enforcing. Eyal Zamir, *The Inverted Hierarchy of Contract Interpretation and Supplementation*, 97 COLUM. L. REV. 1710, 1777–88 (1997).

made when drafting the contract — choices aimed at allocating rights and responsibilities through language.¹³³

When disputes arise, parties frequently assign different meanings to the same contractual language. In some instances, the terms used may conceal divergent understandings, rendering the contract invalid for lack of mutual assent.¹³⁴ In others, unforeseen factual developments or emerging technologies challenge the boundaries of the original language, creating ambiguity about what the parties meant at the time of formation.¹³⁵

In adjudicating such disputes, courts seek the communicative content of the disputed terms — that is, how an ordinary reader would have understood the language at the time of contracting.¹³⁶ In many jurisdictions, this deference to ordinary meaning is codified. For example, California law instructs that “the words of a contract are to be understood in their ordinary and popular sense, rather than according to their strict legal meaning; unless used by the parties in a technical sense, or unless a special meaning is given to them by usage, in which case the latter must be followed.”¹³⁷

Deferring to the ordinary meaning — reflecting the understanding of the general public, including relevant jargon and industry practices¹³⁸ — helps courts reconstruct an objective “meeting of the minds.”¹³⁹ In doing so, they protect reliance interests based on the contract’s apparent meaning, unless there is evidence that the parties shared a different private understanding.¹⁴⁰

133. The objective theory of contract interpretation is well accepted, and it suggests that courts discern objective intent from the words used. E. ALLAN FARNSWORTH, *CONTRACTS* § 3.6 (4th ed. 2004) (discussing prevalence of objective theory); see also Juliet P. Kostritsky, *Interpretive Risk and Contract Interpretation: A Suggested Approach for Maximizing Value*, 2 *ELON L. REV.* 109, 109 (2011) (“The court’s enforcement role begins with ascertaining the parties’ intent objectively by looking to the contract’s language.”).

134. Benjamin Alarie, *Mutual Misunderstanding in Contract*, 46 *AM. BUS. L.J.* 531, 531 (2009) (“Since that time it has been generally accepted in Anglo-American common law that unenforcement is the natural outcome in cases involving mutual misunderstanding.”).

135. See generally Karen Chen, Note, *Catching Up to a New Normal: The Effects of Shifting Industry Standards on Contract Interpretation*, 90 *FORDHAM L. REV.* 2839 (2022) (discussing how courts interpreted contracts impacted by the COVID-19 pandemic).

136. See *supra* note 133.

137. Cal. Civ. Code § 1644 (2024).

138. Including standardized jargon and industry practices. U.C.C. § 1-303(d) (AM. L. INST. & UNIF. L. COMM’N. 2022) (“[A] course of dealing between the parties . . . may give particular meaning to specific terms of the agreement, and may supplement or qualify the terms of the agreement.”).

139. Ben-Shahar & Strahilevitz, *supra* note 114, at 1759–60.

140. Randy E. Barnett, *The Sound of Silence: Default Rules and Contractual Consent*, 78 *VA. L. REV.* 821, 858–59 (1992) (“[T]he purpose for which we adopt the objective approach — to enable persons to rely on the appearances created by others because subjective intentions are generally inaccessible — is satisfied by actual knowledge that the appearances in this case are deceiving. Therefore, in contract law, we protect a party’s reliance on objective appearances, unless it can be shown that the parties shared a common subjective

Importantly, courts do not assess ordinary meaning in a vacuum. Interpretation is shaped by both the textual and commercial contexts of the agreement.¹⁴¹ Courts routinely consider extrinsic factors such as market conditions, course of performance,¹⁴² course of dealing,¹⁴³ and trade usage.¹⁴⁴ In consumer-facing boilerplate contracts, courts may adopt the perspective of a reasonable layperson. By contrast, in negotiated business agreements, they often look to specialized meanings within a particular industry.¹⁴⁵

The *linguistic meaning* of the words is different from their *legal meaning*, namely, the legal significance of particular language in the contract. Courts are not compelled to enforce the ordinary meaning and may decide to advance a different meaning based on interpretive heuristics and legal reasoning.

understanding of the term.”). Clearly, if parties can indicate a mutually agreed upon subjective intent that differs from the objective intent, this intent should prevail. *See* Lawrence M. Solan, *Contract as Agreement*, 83 NOTRE DAME L. REV. 353, 364 (2007) (“[T]he legal system concerns itself principally with the actual intent of the parties and uses ordinary understanding as a surrogate for that intent unless there exists better evidence taken from the transaction itself.”); *see, e.g.*, Taylor v. State Farm Mut. Auto. Ins. Co., 854 P.2d 1134, 1139 (Ariz. 1993) (“If, for example, parties use language that is mutually intended to have a special meaning, and that meaning is proved by credible evidence, a court is obligated to enforce the agreement according to the parties’ intent, even if the language ordinarily might mean something different.”) (internal citation omitted); *see also* TKO Equip. Co. v. C & G Coal Co., 863 F.2d 541, 545 (7th Cir. 1988) (“Under the prevailing will theory of contract, parties, like Humpty Dumpty, may use words as they please. If they wish the symbols ‘one Caterpillar D9G tractor’ to mean ‘500 railroad cars full of watermelons,’ that’s fine — provided parties share this weird meaning.”).

141. *See* Silverstein, *infra* note 171, at 97–98.

142. U.C.C. § 1-303(a) (AM. L. INST. & UNIF. L. COMM’N. 2022) (“A ‘course of performance’ is a sequence of conduct between the parties to a particular transaction that exists if: (1) the agreement of the parties with respect to the transaction involves repeated occasions for performance by a party; and (2) the other party, with knowledge of the nature of the performance and opportunity for objection to it, accepts the performance or acquiesces in it without objection.”).

143. *Id.* § 1-303(b) (“A ‘course of dealing’ is a sequence of conduct concerning previous transactions between the parties to a particular transaction that is fairly to be regarded as establishing a common basis of understanding for interpreting their expressions and other conduct.”).

144. *Id.* § 1-303(c) (“A ‘usage of trade’ is any practice or method of dealing having such regularity of observance in a place, vocation, or trade as to justify an expectation that it will be observed with respect to the transaction in question. The existence and scope of such a usage must be proved as facts. If it is established that such a usage is embodied in a trade code or similar record, the interpretation of the record is a question of law.”); *id.* § 1-303(d). For discussion, *see* Silverstein, *infra* note 171, at 96; JOSEPH M. PERILLO, CALAMARI & PERILLO ON CONTRACTS § 3.9 (6th ed. 2009); BURTON, *supra* note 132, at 41.

145. Stephen J. Choi & G. Mitu Gulati, *Contract as Statute*, 104 MICH. L. REV. 1129, 1131 (2006). (“In discerning this intent, the court may need to look to the overall history of a term, the process by which the term became a standard (or one of the standards) in the industry, and its context within the greater commercial environment.”); Richard A. Posner, *The Law and Economics of Contract Interpretation*, 83 TEX. L. REV. 1581, 1588 (2005) (noting that “[h]istorically, Anglo-American contract law derived from the law merchant, the set of customary norms created by businessmen; such norms would carry a presumption of efficiency”).

Ordinary meaning also plays a central role in applying interpretive heuristics, such as the *plain meaning rule*, which bars extrinsic evidence when contractual language appears clear and unambiguous.¹⁴⁶ But to determine whether language is ambiguous, courts must first engage in the very inquiry into ordinary meaning that the rule presupposes.¹⁴⁷ How should courts determine such ambiguity? As many scholars have noted, U.S. contract interpretation law reflects conflicting approaches, each aimed at advancing different policy goals and relying on various canons of interpretation.¹⁴⁸

Two dominant approaches shape the determination of contractual meaning: *textualism* and *contextualism*.¹⁴⁹ Textualism emphasizes the four corners of the contract, using dictionaries, grammar, and syntax to assess whether the disputed terms have a plain or ambiguous meaning.¹⁵⁰ It prioritizes the public, objective meaning of the chosen words.¹⁵¹ Deference to textual meaning is often justified on the grounds that it promotes legal certainty by aligning with the parties' expectations, discourages strategic behavior by foreclosing post hoc claims about actual intent, and enhances efficiency by limiting the range of interpretive sources courts must consider.

By contrast, contextualism permits the use of extrinsic evidence — including prior negotiations and customary industry practices — *before* determining whether ambiguity exists.¹⁵² In practice, however, both schools converge at the second stage: once a court finds ambiguity, it

146. Margaret N. Kniffin, *Conflating and Confusing Contract Interpretation and the Parol Evidence Rule: Is the Emperor Wearing Someone Else's Clothes?*, 62 RUTGERS L. REV. 75, 95 (2009).

147. See Omri Ben-Shahar, *Data Driven Contract Interpretation: Discovering "Plain Meaning" Through Quantitative Methods*, JOTWELL (June 13, 2018), <https://contracts.jotwell.com/data-driven-contract-interpretation-discovering-plain-meaning-through-quantitative-methods/> [<https://perma.cc/CAN9-EF99>].

148. See Ben-Shahar & Strahilevitz, *supra* note 114, at 1757; FARNSWORTH, *supra* note 114, at 439–41; Keith A. Rowley, *Contract Construction and Interpretation: From the "Four Corners" to Parol Evidence (and Everything in Between)*, 69 MISS. L.J. 73, 82 (1999); Alexander Volokh, *Choosing Interpretive Methods: A Positive Theory of Judges and Everyone Else*, 83 N.Y.U. L. REV. 769, 773–74 (2008).

149. Silverstein, *infra* note 171, at 95–96; Ronald J. Gilson, Charles F. Sabel & Robert E. Scott, *Text and Context: Contract Interpretation as Contract Design*, 100 CORN. L. REV. 23, 25–26 (2014); Shahar Lifshitz & Elad Finkelstein, *A Hermeneutic Perspective on the Interpretation of Contracts*, 54 AM. BUS. L.J. 519, 520 (2017).

150. *Textualism*, LEGAL INFO. INST., <https://www.law.cornell.edu/wex/textualism> [<https://perma.cc/9RF5-MDSF>].

151. Joshua M. Silverstein, *Contract Interpretation Enforcement Costs: An Empirical Study of Textualism Versus Contextualism Conducted Via the West Key Number System*, 47 HOFSTRA L. REV. 1011, 1014 (2019) (“‘Textualist’ judges and commentators argue that the interpretation of contracts should focus primarily on the language contained within the four corners of written agreements.”).

152. Gilson, Sabel & Scott, *supra* note 149, at 38.

typically admits extrinsic evidence to resolve it.¹⁵³ In both cases, a threshold question for contract interpretation starts with a search for the ordinary meaning, in order to determine whether the contractual text encompasses the intended meanings claimed by the litigants.

Both interpretive approaches face well-known critiques. Textualism is often criticized for being overly formalistic and inattentive to commercial realities.¹⁵⁴ By privileging ostensibly objective definitions, textualist methods may obscure the parties' actual intent.¹⁵⁵ Contextualism, while more flexible, has been faulted for inviting uncertainty and reducing predictability — two features that undermine the rule-of-law values contract law purports to serve.¹⁵⁶

Here, majoritarian linguistic signals derived from LLMs can serve as a productive bridge. For textualists, such signals may offer empirically grounded insights into prevailing public usage, strengthening the link between adjudication and ordinary meaning. For contextualists, they may introduce greater consistency and reproducibility into interpretive practices traditionally reliant on case-by-case evidentiary judgment.

At first glance, the use of LLMs to identify ordinary meaning appears well-aligned with the central aims of contract interpretation. LLMs are designed to predict the most statistically likely — or “ordinary” — uses of language in context. However, the most *frequent* usage is not always the most *normatively appropriate*. Standardized boilerplate clauses, for example, while statistically dominant, often reflect the preferences and expectations of drafters — not of the non-drafting parties, especially consumers.¹⁵⁷ Economic analysis has long shown that

153. Silverstein, *infra* note 171, at 100; BURTON, *supra* note 132, at 14; *see, e.g.*, Bank of N.Y. Tr. Co. v. Franklin Advisers, Inc., 726 F.3d 269, 276 (2d Cir. 2013) (applying New York law) (textualist decision); Wagner v. Columbia Pictures Indus., Inc., 52 Cal. Rptr. 3d 898, 901 (Cal. Ct. App. 2007) (contextualist decision).

154. *See supra* note 101.

155. Bradley C. Karkkainen, “Plain Meaning”: Justice Scalia’s Jurisprudence of Strict Statutory Construction, 17 HARV. J.L. & PUB. POL’Y 401, 407 (1994).

156. *See* ANTONIN SCALIA & BRYAN A. GARNER, READING LAW: THE INTERPRETATION OF LEGAL TEXTS 56–66 (2012); *see also* Schwartz & Scott, *supra* note 132, at 547–48 (arguing in favor of textualist interpretation and narrow evidentiary bases that preserve predictability).

157. *See* Steven P. Croley & Jon D. Hanson, *Rescuing the Revolution: The Revived Case for Enterprise Liability*, 91 MICH. L. REV. 683, 770 (1993) (“[T]he efficiency of consumer product markets depends upon consumers’ ability to overcome information costs, for without full information consumers are unable to make consumption and warranty decisions that reflect their true preferences.”); Michael I. Meyerson, *The Efficient Consumer Form Contract: Law and Economics Meets the Real World*, 24 GA. L. REV. 583, 585 (1990) (“[I]mperfect consumer information causes a tendency toward inefficiency in transactions involving consumer form contracts.”); David Gilo & Ariel Porat, *The Hidden Roles of Boilerplate and Standard Form Contracts: Strategic Imposition of Transaction Costs, Segmentation of Consumers and Anticompetitive Effects*, 104 MICH. L. REV. 983, 985 (2006) (similar); *see also* Russell Korobkin, *Bounded Rationality, Standard Form Contracts, and Unconscionability*,

firms drafting such contracts often have incentives to include one-sided or inefficient terms that consumers neither expect nor meaningfully consent to.¹⁵⁸ Thus, in interpreting such contracts, it is essential to ensure that the majoritarian signal extracted from LLMs does not merely mirror the most common usage of a term in boilerplate contracts but instead captures the true expectations of the consumers who are subject to these contracts.¹⁵⁹

Whatever their weight or significance, majoritarian linguistic patterns — if accurate — are likely to play an important role in courts’ interpretive analysis. Thus, the central challenge of incorporating LLM signals in contract interpretation is more methodological than normative.¹⁶⁰ Although current methodologies may not yet resolve all interpretive disputes, they offer a valuable supplement to traditional tools of construction and signal a promising direction for doctrinal development.¹⁶¹

2. Methodology

Traditionally, American courts have relied on dictionary definitions to determine the ordinary meaning of words in contracts.¹⁶² In recent years, legal scholars have increasingly advocated for incorporating empirical evidence into contract interpretation to better align legal outcomes with public linguistic expectations. Empirical methods, they argue, are essential to designing fairer and more effective interpretive rules.¹⁶³

Several empirical approaches have already been applied to contract interpretation, including surveys¹⁶⁴ and corpus linguistics — the study

70 U. CHI. L. REV. 1203, 1204 (2003) (“[Boilerplate contract] forms are often referred to as ‘contracts of adhesion,’ as one party presents the terms to the other on a take-it-or-leave-it basis with no opportunity for negotiation.”); Omri Ben-Shahar, *The Myth of the ‘Opportunity to Read’ in Contract Law*, 5 EUR. REV. CONT. L. 1, 8 (2009) (“[I]t is not reasonable to impose a duty to read the long boilerplate.”); Heather Daiza, *Wrap Contracts: How They Can Work Better for Businesses and Consumers*, 54 CAL. W. L. REV. 201, 211–12 (2017) (explaining that the duty to read becomes “conceptually unfair” when the contracts are incomprehensible). Moreover, this dynamic is exacerbated in the digital sphere, where drafting parties (e.g., internet platforms) have access to extensive data about non-drafting parties (such as platform users). This data allows drafters to identify non-expected terms that are most susceptible to exploitation. See, e.g., Brett Frischmann & Moshe Y. Vardi, *Better Digital Contracts with Prosocial Friction-in-Design*, 65 JURIMETRICS J. 1, 29 (2025).

158. See *supra* note 157.

159. See *infra* note 252 and accompanying text.

160. See Grimmelmann et al., *supra* note 11, at 266–67; Choi, *Unreliable Judges*, *supra* note 11, at 41–43.

161. See Choi, *Unreliable Judges*, *supra* note 11, at 29 (noting that LLMs can reveal “linguistic patterns that bear on interpretive questions”).

162. Kevin P. Tobia, *Testing Ordinary Meaning*, 134 HARV. L. REV. 726, 727 (2020).

163. See, e.g., Ben-Shahar & Strahilevitz, *supra* note 114, at 1756; Stephen C. Mouritsen, *Contract Interpretation with Corpus Linguistics*, 94 WASH. L. REV. 1337, 1360 (2019).

164. See, e.g., Ben-Shahar & Strahilevitz, *supra* note 114, at 1756.

of language based on large, curated collections of texts (corpora).¹⁶⁵ These methods, however, often yield divergent results, exposing important methodological gaps.¹⁶⁶ LLMs, which analyze language usage at scale, offer a more dynamic, efficient, and potentially robust alternative for identifying prevailing linguistic conventions. This Section evaluates key methodologies, focusing on the use of LLMs.

a. Prompting Using Chatbots

In his concurrence in *Snell*, Judge Newsom illustrated how prompting commercial LLM chatbots can be used to extract a majoritarian signal.¹⁶⁷ The judge queried ChatGPT to determine whether installing an in-ground trampoline constituted “landscaping” under a commercial general liability policy. ChatGPT and Bard were both asked two questions: “What is the ordinary meaning of ‘landscaping’?” and “Is installing an in-ground trampoline ‘landscaping’?”¹⁶⁸

According to Judge Newsom, both models gave similar answers to the first question, and ChatGPT’s response aligned with his intuitive sense of how “ordinary people might well use the term ‘landscaping.’”¹⁶⁹ He concluded that commercial LLMs could assist in “ordinary-meaning analyses of legal instruments.”¹⁷⁰ At the same time, however, Judge Newsom also acknowledged significant risks. He noted the potential for hallucinations (fabricated or false information), adversarial prompting (manipulating queries to yield favorable responses), model-shopping, and bias due to population underrepresentation, especially among groups with limited internet presence.¹⁷¹

Judge Newsom’s cautious approach was well placed. Prompting LLMs for ordinary meaning analysis is intuitively appealing but

165. See, e.g., Choi, *Measuring Clarity*, *supra* note 10, at 15.

166. See Tobia, *supra* note 162, at 794–98.

167. *Snell v. United Specialty Ins. Co.*, 102 F.4th 1208, 1221 (11th Cir. 2024).

168. *Id.* at 1224–25.

169. *Id.* at 1225.

170. *Id.* at 1226. This textual analysis was not part of the operative part of the decision because, as Judge Newsom recognized, the immediate case could be resolved without it. Nevertheless, Judge Newsom went to great lengths explaining why LLMs have untapped potential to inform the ordinary-meaning analysis.

171. *Id.* at 1230–32. Judge Newsom noted that “people living in poorer communities (perhaps disproportionately minorities and those in rural areas) are less likely to have ready internet access and thus may be less likely to contribute to the sources from which LLMs draw in crafting their responses to queries.” *Id.* at 1231. Nonetheless, the English-language training bias common to LLMs is less problematic in U.S. courts, which explicitly privilege English-language usage norms in contract interpretation. Joshua M. Silverstein, *Contract Interpretation and the Parol Evidence Rule: Toward Conceptual Clarification*, 24 CHAP. L. REV. 89, 110–11 (2020); RESTATEMENT (SECOND) OF CONTRACTS § 202 cmt. e (AM. L. INST. 1981) (“In the United States the English language is used far more often in a sense which would be generally understood throughout the country than in a sense peculiar to some locality or group. In the absence of some contrary indication, therefore, English words are read as having the meaning given them by general usage, if there is one.”).

methodologically fraught. The practice has sparked a vigorous scholarly debate recently encompassing both methodological reliability and procedural legitimacy.¹⁷² Chatbot outputs are highly sensitive to differences in model architecture, training data, post-training alignment procedures, and prompt design.¹⁷³

In a comprehensive empirical study, Choi demonstrated that off-the-shelf LLMs are highly sensitive to prompt wording, exhibit wide variability across models, and are influenced by post-training interventions such as RLHF.¹⁷⁴ Similarly, Tobia (and later Engel and McAdams) showed that ChatGPT’s interpretive responses shift dramatically with minor phrasing changes — highlighting both the risk of adversarial prompting and the illusion of neutrality.¹⁷⁵

Advocates of corpus linguistics, such as Lee and Egbert, argue that legal interpretation should remain anchored in systematic analysis of curated linguistic data.¹⁷⁶ Compared to LLMs, they contend, corpus methods are transparent, replicable, and capable of distinguishing between different interpretive dimensions — such as conceptual versus applicative meaning. LLMs, by contrast, simulate linguistic plausibility without producing verifiable evidence of actual usage.¹⁷⁷

A key critique of LLM prompting is that its outputs are not replicable for lay users — which, for many scholars, disqualifies them from being treated as genuine empirical evidence. While this objection is valid in strict methodological terms, it may overlook the interpretive value of such outputs in practice. Consider, for example, Judge Newsom’s concurrence in *United States v. Deleon*,¹⁷⁸ where he repeated the same prompt — “What is the ordinary meaning of ‘physically restrained’?” — ten times across three different models (ChatGPT, Claude, and Gemini), generating thirty responses.¹⁷⁹ While the answers were not identical, he noted a shared conceptual core: “the application of tangible force, either through direct bodily contact or some other device or instrument.”¹⁸⁰

172. See Brandon Waldon, Nathan Schneider, Ethan Wilcox, Amir Zeldes & Kevin Tobia, *Large Language Models for Legal Interpretation? Don’t Take Their Word for It*, 114 GEO. L.J. 115, 137–53 (2025); Thomas Lee & Jesse Egbert, *Artificial Meaning?*, 77 FLA. L. REV. (forthcoming 2026) (manuscript at 4), <https://papers.ssrn.com/abstract=4973483> [<https://perma.cc/42P3-UHDW>]; Choi, *Unreliable Judges*, *supra* note 11, at 26.

173. See *supra* Section II.B.2.

174. Choi, *Unreliable Judges*, *supra* note 11, at 8–9.

175. Kevin P. Tobia, *Testing Ordinary Meaning*, 134 HARV. L. REV. 726, 794–98 (2020); Engel & McAdams, *supra* note 10, at 237–40.

176. See Lee & Egbert, *supra* note 172, at 5–6.

177. *Id.*

178. 116 F.4th 1260 (11th Cir. 2024) (Newsom, J., concurring).

179. *Id.* at 1272 (turning to multiple AI tools to interpret the phrase “physically restrained” under the federal sentencing guidelines).

180. *Id.* at 1275.

Far from seeing this variation as a flaw, Judge Newsom framed it as a meaningful insight into language itself. As he explained:

Remember, our aim is to discern ‘ordinary meaning.’ Presumably, the *ideal* gauge of a word’s or phrase’s ordinary meaning would be a broad-based survey of every living speaker of American English — totally unrealistic, but great if you could pull it off. Imagine how that experiment would go: If you walked out onto the street and asked all umpteen million subjects, “What is the ordinary meaning of ‘physically restrained’?”, I think I can confidently guarantee that you would not get the exact same answer spit back at you verbatim over and over and over. Instead, you’d likely get a variety of responses that differed around the margins but that, when considered en masse, revealed a common core. And that common core, to my way of thinking, is the *ordinary* meaning.¹⁸¹

This analysis underscores a distinctive strength of prompting: its ability to surface both the common core and the fuzziness of natural language.¹⁸² While LLM outputs may not be identical or fully replicable, they can reflect real-world semantic variation while still converging on broadly shared meaning. In fact, this variation may help courts determine when a contract term is ambiguous, based on the presence of multiple reasonable interpretations.

Returning to the normative objective of contract interpretation — identifying the meaning mutually intended by the parties — this capacity to capture both variety and convergence provides valuable insight. Rather than undermining judicial reliability, this linguistic texture reflects the actual diversity of language use and helps courts ground their decisions in socially plausible understandings of disputed terms.

Still, courts must proceed cautiously. Scholars have proposed safeguards, including varying prompt phrasing, comparing results across models, and assessing output consistency before drawing inferences.¹⁸³ While Judge Newsom employed several of these techniques in his experiment, broader sampling and more systematic consistency analysis would have strengthened his conclusions.

181. *Id.* at 1276.

182. *But see* limitations discussed *supra* Section II.B.2.

183. Choi, *Unreliable Judges*, *supra* note 11, at 28. *But see* Grimmelmann et al., *supra* note 11, at 258 (casting serious doubt on the viability of such mitigating methods altogether).

b. Moving Beyond Chatbots: Cosine Similarity

Alternative methods for extracting majoritarian linguistic signals are being actively developed. One such method, proposed by Jonathan Choi, relies on cosine similarity between word embeddings.¹⁸⁴ Word embeddings are high-dimensional vector representations that encode semantic information based on word co-occurrence in large textual corpora.¹⁸⁵ Cosine similarity measures the angle between two vectors in this space, providing a quantitative estimate of their semantic closeness.¹⁸⁶ To illustrate how cosine similarity might inform ordinary meaning analysis, Choi adapted H.L.A. Hart’s canonical “no vehicle in the park” hypothetical — a thought experiment designed to demonstrate how even apparently clear legal rules have a penumbra of uncertain meanings. To that end, Choi generated a ranked list of terms based on their proximity to “vehicle” in vector space — ranging from semantically close terms like “car” to more distant ones like “crutches.”¹⁸⁷

Theoretically, a court — such as the one in *Snell* — could use Choi’s method to assess whether “in-ground trampoline installation” is semantically proximate to “landscaping.”¹⁸⁸ If the cosine similarity between the two is high (as with “vehicle” and “car”), this would support the court’s interpretation.¹⁸⁹ In practice, however, the method faces limitations. Multi-word embeddings (e.g., “in-ground trampoline installation”) are less reliable, and isolated keywords (e.g., “trampoline”) often lack sufficient context. These constraints can lead to unstable or misleading similarity metrics.¹⁹⁰

Choi’s method is, therefore, especially useful mainly for simple interpretive exercises involving single-word comparisons. For example, he demonstrated that the word “tomato” is closer in meaning to “vegetable” (cosine similarity = 0.57) than to “fruit” (cosine similarity =

184. Choi, *Measuring Clarity*, *supra* note 10, at 21–23; *see also* Suresh & Gutttag, *supra* note 32, at 5 (“Recent research has shown that word embeddings, which are learned from large corpora of text (e.g., Google news, web pages, Wikipedia), reflect human biases . . . and that an embedding model trained on data from a particular decade reflects the biases of that time.”).

185. Choi, *Measuring Clarity*, *supra* note 10, at 20; Tomas Mikolov, Kai Chen, Greg Corrado & Jeffrey Dean, *Efficient Estimation of Word Representations in Vector Space*, 2013 PROC. INT’L CONF. ON LEARNING REPRESENTATIONS, Jan. 2013, at 1, 2–4 (introducing model architectures for computing word embeddings); Gideon Lewis-Kraus, *The Great A.I. Awakening*, N.Y. TIMES MAG. (Dec. 14, 2016), <https://www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html> [<https://perma.cc/ku6a-ssxt>].

186. Choi, *Measuring Clarity*, *supra* note 10, at 20–22.

187. *Id.* at 24–25, 30–31.

188. *Cf.* Arbel & Hoffman, *supra* note 10, at 454–56 (using this approach).

189. Choi, *Measuring Clarity*, *supra* note 10, at 37–43 (checking real-world cases).

190. *Id.* at 50–59 (discussing limitations related to word meanings and multiple word choices).

0.543) — an interpretive distinction relevant in *Nix v. Hedden* for tariff classification.¹⁹¹

Despite its limitations, cosine similarity offers a low-noise method for extracting majoritarian signals without relying on prompts.¹⁹² Its strength lies in minimizing the variability introduced by prompt design. However, because many interpretive disputes hinge on contextual nuance, prompt-based methods may remain more useful for contractual analysis, where the meaning of terms often depends on surrounding text.

c. Contextual Prompting and Autocompletion

To ascertain the ordinary meaning of the term “landscaping,” Judge Newsom in *Snell* relied on ostensibly context-free queries such as, “What is the ordinary meaning of ‘landscaping?’”¹⁹³ Yet, there is no such thing as a contextless prompt.¹⁹⁴ Lexical choices, and perhaps even punctuation marks, introduce contextual cues that can shape the model’s predictive output.¹⁹⁵ Rather than treating this feature as a flaw, however, the context sensitivity of prompt-driven methods can be leveraged as an asset. By incorporating linguistic context drawn from the contract itself or its surroundings, such as specific clauses, the agreement as a whole, or even related contractual correspondence, majoritarian signals can be rendered more context-aware. In this way, these

191. *Id.* at 40–42 (citing *Nix v. Hedden*, 149 U.S. 304, 306 (1893)).

192. Cosine similarity measurements are a promising methodology for extracting majoritarian signals from LLMs that go beyond contract interpretation. They may assist courts in appraising many legal texts, including evaluating statutes or assessing patent claims for indefiniteness. *See generally* Choi, *Measuring Clarity*, *supra* note 10, at 13, 60.

193. *Snell v. United Specialty Ins. Co.*, 102 F.4th 1208, 1221, 1224–25 (11th Cir. 2024).

194. Pranab Sahoo, Ayush Kumar Singh, Sriparna Saha, Vinija Jain, Samrat Mondal & Aman Chadha, A Systematic Survey of Prompt Engineering in Large Language Models: Techniques and Applications 1 (Mar. 16, 2025) (unpublished manuscript) (on file with arXiv), <https://arxiv.org/pdf/2402.07927v2> [<https://perma.cc/LY9R-L9K2>] (“Prompts can be natural language instructions that provide context to guide the model or learned vector representations that activate relevant knowledge.”); Stefan Hackmann, Haniyeh Mahmoudian, Mark Steadman & Michael Schmidt, Word Importance Explains How Prompts Affect Language Model Outputs 1 (Mar. 6, 2024) (unpublished manuscript) (on file with arXiv), <https://arxiv.org/pdf/2403.03028> [<https://perma.cc/X9C4-6TU4>] (“[V]arying individual words in prompts... uncover[s] their statistical impact on the model outputs.”); Om Dobariya & Akhil Kumar, Mind Your Tone: Investigating How Prompt Politeness Affects LLM Accuracy 1 (unpublished manuscript) (Oct. 6, 2025) (on file with arXiv), <https://arxiv.org/pdf/2510.04950> [<https://perma.cc/LH5V-7JG3>] (“The wording of natural language prompts has been shown to influence . . . performance.”).

195. *See* Grimmelmann et al., *supra* note 11, at 251; Choi, *Unreliable Judges*, *supra* note 11, at 21; Choi, *Measuring Clarity*, *supra* note 10 at 59 (explaining that LLMs “require a user to specify the context in which words should be read, which might allow an interpreter to manipulate the interpretive process.”). *But see* Arbel & Hoffman, *supra* note 10, at 490–91 (experimenting with numerous types of prompts, finding similar outputs despite word changes, but acknowledging the limitations of their methodology).

methods can mediate between textualist and contextualist approaches rather than collapse into either one.¹⁹⁶

Indeed, by using the contract’s original language as a prompt, courts can extract outputs that remain anchored in the “four corners” of the agreement while still drawing from broader usage patterns.¹⁹⁷ This hybrid strategy offers a middle ground: the predictability of textualism combined with the richness of contextualism.

One technique that mitigates many of these problems is to avoid prompt engineering and rely instead on autocompletion (known in computer science jargon as “causally masking”). This method provides the model with surrounding contractual language while omitting the disputed term, prompting the LLM to complete the clause based on its training data. Put simply, the model is shown the contract with a key word hidden and is asked to fill in the blank the way most writers typically would. Because it leverages the actual contractual language — rather than external or synthetic prompts — autocompletion reduces the introduction of irrelevant noise and yields more targeted insights into prevailing usage patterns.

Autocompletion is particularly helpful for interpreting vague or open-textured terms, such as “reasonable time,” which are not easily clarified by dictionaries or traditional canons.¹⁹⁸ Consider *Maple Manor Rehabilitation Center v. Evanston Insurance Co.*,¹⁹⁹ where the court evaluated whether a two-year delay in reporting a claim violated a policy’s requirement to file “as soon as practicable.”²⁰⁰ Without citing any empirical basis, the court deemed the delay unreasonable.²⁰¹

To test whether LLMs could assist in this determination, we ran an experiment to determine how an autocompletion inquiry could have aided the court’s analysis in the above case. This serves as an illustrative example of the types of results courts could receive from using publicly available, unmodified models, and is limited by the

196. See Arbel & Hoffman, *supra* note 10, at 510.

197. See *id.* at 451.

198. Traditional empirical tools such as surveys can help the courts in such interpretative cases. However, these tools are far more expensive and also subject to serious methodological limitations. See *supra* note 102.

199. No. 359147, 2023 WL 3131930 (Mich. Ct. App. Apr. 27, 2023) [hereinafter *Maple*].

200. *Id.* at 6–7 (“The phrase ‘as soon as practicable’ has, without exception, assumed a working definition with a very specific legal understanding — that of being within a reasonable time under the circumstances.”).

201. The court concluded that “nearly two years is not a reasonable amount of time.” *Id.* at 7. (“Nearly two years is not a reasonable amount of time in any circumstance where Maple Manor fully participated in the lawsuit with the same attorneys it has now and has offered no justifiable excuse and identified no impediment to informing defendant of the Irvine lawsuit prior. Maple Manor has offered no substantive argument or caselaw suggesting that it is and we find none.”).

shortcomings of those models. At the time of this writing, the authors are conducting a fuller validating experiment using a fine-tuned model.²⁰²

For this early test, we masked the phrase “as soon as practicable” in the policy clause and prompted GPT-2 to autocomplete the sentence.²⁰³ The model generated: “[a] claim shall be filed with the insurance company within 30 days after the date on which the claim is filed.” Repeated runs produced the same temporal estimate.²⁰⁴

Even without extracting formal probability distributions or validating results across multiple models, this basic experiment revealed two key patterns: (1) reporting periods are likely to be expressed in numeric terms (e.g., days, weeks), and (2) such periods rarely extend to years.²⁰⁵

202. In this ongoing project, we examine whether standard consumer-facing LLMs reproduce real-world majoritarian patterns previously documented in the scholarly literature. For example, in the privacy context, Florencia Marotta-Wurgler and David Stein have assembled a comprehensive, expert-labeled dataset of standard-form contractual provisions. See Florencia Marotta-Wurgler & David Stein, *Building a Long Text Privacy Policy Corpus with Multi-Class Labels*, 1 PROC. 63RD ANN. MEETING ASS’N FOR COMPUTATIONAL LINGUISTICS 8156 (2025). That dataset provides a useful benchmark for testing whether the autocomplete responses of consumer-facing generative AI models reflect prevailing market norms. We are currently evaluating this question using their dataset.

203. The model we used is GPT-2 (the smallest version (124M parameters) obtained from *Hugging Face*). We trained the model to generate text given an input. The input text was tokenized using the GPT-2 tokenizer. We took as input the beginning of Section 4 from an insurance policy from *Maple* explicitly: “A. Claim Reporting Provision: It is a condition precedent to coverage afforded by this Coverage Part that the Insured shall give to the Company written notice as stated in the Notices item of the Declarations *as soon as practicable*.” *Id.* at 6 (emphasis added). We then masked the tilted term and let the model complete the missing gap. The model’s temperature was set to one (the default value). The maximum length of the generated output was 500 tokens. We conducted three iterations.

204. The model generated the same text every time. Notably, other than using the thirty-day time limit, the model’s autocompleted text was largely illogical. Specifically, the auto-completed term began with the following:

A claim shall be filed with the insurance company within 30 days after the date on which the claim is filed. If the Claim is not filed within the 30-day period specified in subparagraph (A), the insurer shall file a claim within 60 days of receipt of such claim. In the event of a delay in filing a Claim, the claimant shall not be entitled to a refund of any amount paid to him or her in connection with such delay.

205. Otherwise, the model’s top autocompletion suggestion was also “as soon as practicable.” Even in such cases, however, courts may acquire valuable signaling information by deferring to the other less probable terms in the autocompletion hierarchy. Note also that if the model’s top suggestion for “as soon as practicable” was “as soon as practicable,” it would have been difficult for the court to know whether the model’s suggestion reflects a genuine majoritarian signal in relevant data or was merely a glitch where the model has “memorized” specific phrases from its training set. Cf. Myra Cheng, Tiziano Piccardi & Diyi Yang, CoM-PosT: Characterizing and Evaluating Caricature in LLM Simulations 1 (Oct. 12, 2023) (unpublished manuscript) (on file with arXiv), <https://arxiv.org/abs/2310.11501> [<https://perma.cc/KD2W-PDK9>] (discussing this problem). However, the converse is also true. If the model suggests a meaningfully different term from the masked term — such as “30 days” instead of “as soon as practicable” in the *Maple* case — this signal may have genuine empirical value. Such results indicate that the masked term is relatively uncommon in

These observations could support a court’s finding that a two-year delay is inconsistent with ordinary usage.²⁰⁶

Of course, this evidence is not dispositive. A court may choose to follow the majoritarian signal to reinforce its conclusion, or it may find that the vague term was intentionally chosen and that the signal is in-apposite.²⁰⁷ Judicial interpretation ultimately requires normative judgment. Courts may weigh LLM-derived signals differently depending on factors such as party sophistication, context of formation, or bargaining asymmetries, particularly in contracts of adhesion like insurance policies.²⁰⁸

d. Domain-Specific Modeling and Future Methodologies

Interpretive methodologies grounded in LLMs are likely to evolve rapidly, driven by innovations in signal extraction,²⁰⁹ domain-specific fine-tuning, and increased judicial validation and familiarity.²¹⁰ Consider the recent case *Pegnatori v. Pure Sports Technologies*,²¹¹ which involved interpreting the term “foam” in a patent claim for softball bat technology.

Unlike contract disputes, patent claim construction is governed by strict temporal and contextual constraints. Courts must interpret terms based on their meaning at the time of filing, from the perspective of a

contracts similar to the one provided to the model. This is because the rich context provided to the model — typically the entire contract or substantial excerpts — encourages the model to rely on memorization of known textual patterns. Thus, if the model nevertheless generates a different term despite a strong contextual prompt pushing it to memorize, it suggests that the original masked term was relatively uncommon or ambiguous.

206. A more rigorous approach would be to use the entire contract as the relevant context and use many more models. In addition, because GPT-2 couldn’t handle in-test masking we had to let the model complete the entire paragraph while defining for it a limited length. A better approach would be to use the context before as well as after the masked term.

207. See *supra* Sections II.B.1., III.A.1.

208. See *infra* Section III.A.3.c.; Margaret N. Kniffin, CORBIN ON CONTRACTS: INTERPRETATION OF CONTRACTS § 24.27 (Joseph M. Perillo ed., rev. ed. 1998); see also Michelle Boardman, *Penalty Default Rules in Insurance Law*, 40 FLA. ST. U. L. REV. 305, 305 (2013); Horton, *supra* note 119, at 431; Boardman, *supra* note 119, at 1105; Klapp v. United Ins. Grp. Agency, Inc., 663 N.W.2d 447, 454 (Mich. 2003) (“In interpreting a contract whose language is ambiguous . . . ambiguities are to be construed against the drafter of the contract.”); Michael B. Rappaport, *The Ambiguity Rule and Insurance Law: Why Insurance Contracts Should Not Be Construed Against the Drafter*, 30 GA. L. REV. 171, 254 (1995); Bradley D. Liggett, *Contra Applicantem or Contra Proferentem Applicatio: The Need For Clarification of the Doctrine of Contra Proferentem in the Context of Insured-Created Ambiguities in Insurance Applications*, 2008 BYU L. REV. 211, 221–22 (2008); Ethan J. Leib & Steve Thel, *Contra Proferentem and the Role of the Jury in Contract Interpretation*, 87 TEMP. L. REV. 773, 790 (2015); Kenneth S. Abraham, *A Theory of Insurance Policy Interpretation*, 95 MICH. L. REV. 531, 531 (1996).

209. See, e.g., Hacothen & Elkin-Koren, *supra* note 10, at 555–56.

210. For example, specialized ordinary meaning datasets could be developed to facilitate such interpretive exercises. Choi, *Unreliable Judges*, *supra* note 11, at 28.

211. No. 2:23-cv-01424, 2023 WL 6626159, at *5–6 (D.S.C. Oct. 11, 2023).

person of ordinary skill in the art (“POSITA”), and in light of the patent’s prosecution history and technical disclosures.²¹²

In *Pure Sports*, the court declined to rely on ChatGPT’s definition of “foam,” citing the model’s lack of domain specificity and temporal precision.²¹³ However, future litigants may introduce fine-tuned LLMs, trained on domain-specific corpora — such as patent disclosures in materials science from 1995–2005 — to simulate the interpretive perspective of a skilled artisan from that period.²¹⁴

These models would introduce new questions, including how to define the relevant “art” or ensure that training data reflects the relevant temporal window. Yet such questions already arise in the context of human expert testimony. In many cases, fine-tuned models may offer greater transparency, reproducibility, and resistance to financial bias than traditional experts.²¹⁵

Ultimately, all interpretive methodologies — whether embedding-based, prompt-driven, or fine-tuned — should be subject to rigorous empirical validation before or during their adoption in litigation.²¹⁶ Procedural safeguards are essential to ensure the admissibility, contestability, and evidentiary reliability of majoritarian signals.²¹⁷ The next subsection turns to those procedural considerations.

3. Legal Procedure

Critics of using LLMs in contract interpretation often argue that LLMs cannot, and should not, resolve cases on their own.²¹⁸ But this critique mischaracterizes the actual proposals advanced by many

212. See, e.g., *Pfizer, Inc. v. Teva Pharms., Inc.*, 429 F.3d 1364, 1372–73 (Fed. Cir. 2005) (citing *Phillips v. AWH Corp.*, 415 F.3d 1303, 1313 (Fed. Cir. 2005) (en banc)); *Z-Man Fishing Prods., Inc. v. Renosky*, 790 F. Supp. 2d 418, 426 (D.S.C. 2011) (citing *Markman v. Westview Instruments, Inc.*, 52 F.3d 967, 979–80 (Fed. Cir. 1995), *aff’d*, 517 U.S. 370 (1996)).

213. 2023 WL 6626159, at *6; cf. Suresh & Gutttag, *supra* note 32, at 5 (explaining the problem of “aggregation bias” where “a one-size-fits-all model is used for data in which there are underlying groups or types of examples that should be considered differently”).

214. See Choi, *Unreliable Judges*, *supra* note 11, at 31.

215. See Roth, *supra* note 100, at 1972 (describing credibility problems and solutions for machine bias). Clearly, to be legitimate, computational evidence should be submitted by human experts who can be accountable for their methods and subject to cross-examination. *Phillips*, 415 F.3d at 1318 (noting that expert testimony is typically procured for the purpose of litigation and therefore can suffer from bias, and “[t]he effect of that bias can be exacerbated if the expert is not one of skill in the relevant art or if the expert’s opinion is offered in a form that is not subject to cross-examination”).

216. See *supra* note 202 and accompanying text.

217. See Choi, *Unreliable Judges*, *supra* note 11, at 28–31; Grimmelmann et al., *supra* note 11, at 297–308.

218. E.g., Choi, *Unreliable Judges*, *supra* note 11, at 30 (concluding that off-the-shelf LLMs cannot explicate real-world interpretive questions).

proponents — including this Article.²¹⁹ Supporters do not suggest that LLM outputs are conclusive or dispositive. Rather, they argue that when derived using sound methodology, LLMs can provide one piece of probative evidence among others in a broader interpretive inquiry. Just as courts consult dictionaries, surveys, and canons, LLMs may offer input to judicial reasoning — but they do not themselves constitute legal conclusions regarding meaning or its normative weight.

Still, incorporating majoritarian signals from LLMs into contract interpretation raises two primary categories of concern: (1) the persuasive but potentially manipulative nature of LLM outputs, and (2) their lack of predictive precision.²²⁰

First, scholars have warned that LLMs can produce persuasive yet epistemically shallow interpretations — outputs that appear objective and credible, even when they are neither.²²¹ Such tools could be strategically manipulated by litigants or judges, especially through adversarial prompting or selective model choice. Critics argue that these outputs may offer the illusion of neutrality while masking normative or methodological biases.²²² However, these risks can be mitigated by integrating LLM signals into an adversarial process, where opposing parties can contest the choice of model, the phrasing of prompts, and the reliability of results. Judges, moreover, are trained to evaluate evidentiary claims critically and to arbitrate among competing interpretations.

Second, critics question whether LLMs can accurately simulate how an “ordinary reader” would understand legal language. Prompt sensitivity, model variance, and alignment tuning (e.g., RLHF) all contribute to unstable and inconsistent outputs.²²³ Jonathan Choi has demonstrated substantial disagreement between different LLMs on interpretive questions, undermining claims to predictive reliability.²²⁴ Moreover, post-training refinement and fine-tuning may distort majoritarian signals to reflect normative goals — such as fairness or safety —

219. Arbel and Hoffman argue that LLMs can assist courts by quantifying ambiguity, predicting majoritarian meaning, and evaluating extrinsic evidence more cheaply and systematically than traditional methods. See Arbel & Hoffman, *supra* note 10, at 451. At the same time, however, they stress that generative interpretation is not a replacement for adjudication but a practical and transparent supplement for a resource-constrained legal system. *Id.*

220. See *supra* Section II.B.2.

221. See Zihao Li, Weiwei Yi & Jiahong Chen, Beyond Accuracy: Rethinking Hallucination and Regulatory Response in Generative AI 1 (Oct. 23, 2025) (unpublished manuscript) (on file with arXiv), <https://arxiv.org/pdf/2509.13345> [<https://perma.cc/2HNB-PFGU>] (“[H]allucinated content in language models may appear fluent, persuasive, and contextually appropriate while conveying distortions that escape conventional accuracy checks.”).

222. Grimmelmann et al., *supra* note 11, at 272 (warning against confusing fluency with validity).

223. Choi, *Unreliable Judges*, *supra* note 11, at 8–10, 21; Grimmelmann et al., *supra* note 11, at 252; Waldon et al., *supra* note 172, at 150–53; see Lee & Egbert, *supra* note 172, at 6 (advocating empirically grounded but more interpretable alternatives like corpus linguistics).

224. Choi, *Unreliable Judges*, *supra* note 11, at 24.

rather than actual usage.²²⁵ For example, when assessing political bias profiles in GenAI models, Faulborn et al. have found “that instruction-tuning significantly shifts the political position of the models to the left when compared to their base version.”²²⁶ Clearly, if the goal of majoritarian signals is to reflect real-world distribution, this form of instruction-tuning bias would make such a signal less effective.

In the subsections that follow, we explore whether legal procedure can address these concerns. We consider how rules distinguishing law from fact, evidentiary admissibility doctrines, and burden-shifting mechanisms might provide safeguards for courts seeking to incorporate LLM-generated signals into contract interpretation.

a. Questions of Law and Fact

Contract interpretation involves both legal and factual questions. According to the textualist approach, judges determine as a matter of law whether a disputed term is ambiguous — that is, whether it is reasonably susceptible to more than one meaning.²²⁷ If ambiguity is found, interpretation often shifts to a factual inquiry, typically resolved at trial with the aid of extrinsic evidence such as negotiation history, course of dealing, and industry norms.²²⁸

Ordinary meaning functions as a legal baseline, while the ambiguity threshold determines whether courts must admit factual evidence to resolve interpretive disputes.²²⁹ When ordinary meaning is treated as a question of law, judges retain authority over interpretation and may consult LLMs as advisory tools — akin to dictionaries or interpretive canons.²³⁰ In this setting, LLM outputs can illuminate how a typical reader might understand a term, while still leaving ultimate interpretive authority with the court.

Critics argue that LLMs cannot produce universal, replicable, or normatively calibrated conclusions about ordinary meaning. But courts do not pursue scientific truth in this context; rather, they seek pragmatic

225. *Id.* at 27.

226. Mats Faulborn, Indira Sen, Max Pellert, Andreas Spitz & David Garcia, Only a Little to the Left: A Theory-grounded Measure of Political Bias in Large Language Models 6 (July 20, 2025) (unpublished manuscript) (on file with arXiv), <https://arxiv.org/abs/2503.16148> [<https://perma.cc/MEX4-UKRW>].

227. Aaron D. Goldstein, *The Public Meaning Rule: Reconciling Meaning, Intent, and Contract Interpretation*, 53 SANTA CLARA L. REV. 73, 80–81 (2013); Peter Linzer, *The Comfort of Certainty: Plain Meaning and the Parol Evidence Rule*, 71 FORDHAM L. REV. 799, 803–06 (2002).

228. Jurisdictions that adhere to the contextualistic approach will consider contextual evidence even before resolving the question of ambiguity. See Aaron D. Goldstein, *The Public Meaning Rule: Reconciling Meaning, Intent, and Contract Interpretation*, 53 SANTA CLARA L. REV. 73, 80 (2013); Ronald J. Gilson, Charles F. Sabel & Robert E. Scott, *Text and Context: Contract Interpretation as Contract Design*, 100 CORN. L. REV. 23, 34–36 (2014).

229. Choi, *Measuring Clarity*, *supra* note 10, at 7.

230. See *id.* at 14.

benchmarks for determining what contracting parties intended or reasonably expected.²³¹ In *Snell*, for example, Judge Newsom queried ChatGPT to assess whether installing a trampoline constituted “landscaping.”²³² His use of the LLM was advisory, not dispositive.

This example highlights a key procedural strength: Professional judges are, in theory, generally less susceptible than juries to manipulation by persuasive but hollow rhetoric. Their training enables them to scrutinize competing arguments, apply evidentiary standards, and cabin the influence of unreliable tools. At the same time, integrating LLM outputs into judicial reasoning may blur the line between legal and factual inquiries. As reliance on these tools grows, questions of meaning could evolve into mixed questions of law and fact — implicating evidentiary standards and procedural safeguards more commonly associated with expert testimony.²³³

This shift underscores the need for clear doctrinal guidance. Courts must develop coherent procedures for admitting, evaluating, and contesting LLM-derived signals in ways that remain faithful to the normative foundations of contract law while adapting to modern technological realities.

b. Lessons From Surveys

In a recent proposal, Ben-Shahar and Strahilevitz argued that the task of interpreting contractual language should be delegated to surveys.²³⁴ Their “survey interpretation method” would resolve disputes through randomized experiments measuring how representative populations understand contested terms.²³⁵ Compared to this sweeping proposal, using LLMs to inform legal reasoning about ordinary meaning is far more modest — and more closely aligned with existing judicial practice.

In *United States v. Deleon*, Judge Newsom explicitly analogized LLM prompting to surveys.²³⁶ Both methods aim to estimate how a

231. See *supra* Section III.A.1.

232. *Snell v. United Specialty Ins. Co.*, 102 F.4th 1208, 1221 (11th Cir. 2024).

233. *Id.* at 1232 (emphasizing that LLMs are but another tool among many to “provide additional datapoints to be used alongside dictionaries, canons, and syntactical context in the assessment of terms’ ordinary meaning”).

234. Ben-Shahar & Strahilevitz, *supra* note 114, at 1824 (arguing that the meaning of disputed contractual clauses should be determined by polling a large representative sample of disinterested respondents).

235. *Id.* at 1758, 1824.

236. Judge Newsom wrote:

Technical explanations aside, why, in retrospect, should I have been so unnerved by the models’ slightly (though again, not substantively) different answers? . . . Remember, our aim is to discern ‘ordinary meaning.’ Presumably, the *ideal* gauge of a word’s or phrase’s ordinary

term is understood by a relevant population. Surveys gather structured responses from human subjects; LLMs generate linguistic completions based on statistical patterns in training data. While surveys benefit from a mature methodological tradition,²³⁷ LLM technology remains more volatile, and best practices are still evolving.

Yet many objections raised against LLMs echo early criticisms of surveys. When first introduced in the early twentieth century, surveys were widely rejected as hearsay — out-of-court statements offered for the truth of the matter asserted.²³⁸ Today, however, surveys are routinely admitted in trademark and unfair competition litigation, particularly to assess consumer confusion.²³⁹ Unlike cases of trademark or false advertising, courts interpreting consumer contracts generally do not determine meaning by surveying representative consumers, instead relying on canons, policy arguments, business-sense conjectures, or dictionaries, although scholars have recently proposed introducing new survey-based methods for resolving contract disputes.²⁴⁰

Surveys remain subject to challenge. Courts scrutinize their methodological soundness, including sampling, questionnaire design, and relevance to the dispute. Flaws such as leading questions, ambiguous

meaning would be a broad-based survey of every living speaker of American English — totally unrealistic, but great if you could pull it off. Imagine how that experiment would go: If you walked out onto the street and asked all umpteen million subjects, “What is the ordinary meaning of ‘physically restrained’?”, I think I can confidently guarantee that you would *not* get the exact same answer spit back at you verbatim over and over and over. Instead, you’d likely get a variety of responses that differed around the margins but that, when considered en masse, revealed a common core. And that common core, to my way of thinking, is the *ordinary* meaning. So, as it turns out, the very thing that had initially given me pause — namely, that the LLMs were returning subtly different responses to the same question — has instead given me (more) hope that the models have something significant to offer the interpretive enterprise. The fact is, language is an organic thing, and like most organic things, it can be a little messy. So too, unsurprisingly, are our efforts to capture its ordinary meaning. Because LLMs are trained on actual individuals’ uses of language in the real world, it makes sense that their outputs would likewise be less than perfectly determinate — in my experience, a little (but just a little) fuzzy around the edges.

116 F.4th 1260, 1276–77 (11th Cir. 2024) (Newsom, J., concurring).

237. See Ioana Vasii & Lucian Vasii, *Survey Evidence in Trademark Actions*, 22 DEPAUL BUS. & COM. L.J. 57, 61–62, 68–71 (2024).

238. *Schering Corp. v. Pfizer, Inc.*, 189 F.3d 218, 224 (2d Cir. 1999) (“In the first half of this century, surveys were generally regarded as inherently untrustworthy because they contained hearsay, or out-of-court statements offered to prove the truth of the matters asserted.”).

239. Vasii & Vasii, *supra* note 237, at 57. However, courts are cautioned to treat them carefully and regard them as merely one component of the multifaceted likelihood of confusion analysis. See *Jack Daniel’s Props., Inc. v. VIP Prods., LLC*, 599 U.S. 140, 163 (2023) (Sotomayor, J., concurring) (“[C]ourts should treat the results of surveys with particular caution . . . Like any other evidence, surveys should be understood as merely one piece of the multifaceted likelihood of confusion analysis.”).

240. *E.g.*, Ben-Shahar & Strahilevitz, *supra* note 114, at 1758.

phrasing, or unrepresentative samples can undermine reliability.²⁴¹ In the trademark context, for instance, surveys conducted over the phone may fail to enable visual inspection of the trademarks and therefore fail to reflect actual marketing conditions.²⁴² Similarly, strategic wording choices of survey questions can affect the legal outcomes by suggesting the “right” answers (leading questions),²⁴³ introducing bias (closed-ended questions)²⁴⁴ or creating confusion through ambiguity.²⁴⁵ As Justice Sotomayor recently cautioned in her concurrence to *Jack Daniel’s Properties, Inc. v. VIP Products, LLC*,²⁴⁶ even “cleverly designed” surveys can induce confusion by prompting artificial legal reasoning in respondents.²⁴⁷

Importantly, most challenges to survey evidence go to weight, not admissibility.²⁴⁸ Courts rarely exclude surveys outright, but adjust how much persuasive value they assign based on methodological rigor.²⁴⁹ This calibrated approach provides a useful model for courts considering LLM outputs. Just as survey flaws do not bar their use, neither should LLMs be categorically excluded. Instead, courts should scrutinize the methodology, weigh limitations, and incentivize refinement.

c. Shifting the Burden of Proof

Burden shifting can serve as an important procedural safeguard when courts rely on LLM-derived majoritarian signals in contract interpretation. Majoritarian outputs may be a useful proxy for ordinary meaning, but they can also reflect dominant drafting practices rather than the parties’ actual understandings, especially in boilerplate

241. See Vasu & Vasu, *supra* note 237, at 71–75; see also Michael J. Stern, Ipek Bilgen, & Don A. Dillman, *The State of Survey Methodology: Challenges, Dilemmas, and New Frontiers in the Era of the Tailored Design*, 26 *FIELD METHODS* 284, 285 (2014) (survey results are also contingent on how the relevant universe is defined and how the survey’s sample is selected).

242. Vasu & Vasu, *supra* note 237, at 68.

243. *Procter & Gamble Co. v. Ultreo, Inc.*, 574 F. Supp. 2d 339, 352 (S.D.N.Y. 2008) (“A survey is not credible if it relies on leading questions which are inherently suggestive and invite guessing by those who did not get any clear message at all.” (internal citation omitted)).

244. *Simpson Strong-Tie Co. v. Mitek Inc.*, No. 20-VC-06957, 2023 WL 137478, at *5 (N.D. Cal. Jan. 9, 2023) (order denying motion to exclude expert witness testimony but cautioning against reliance on close-ended questions).

245. *Jack Daniel’s Props., Inc. v. VIP Prods., LLC*, 599 U.S. 140, 163–65 (2023) (Sotomayor, J., concurring).

246. 599 U.S. 140 (2023).

247. In this case, VIP Products produced a dog toy (“Bad Spaniels”) that parodied Jack Daniel’s whiskey bottle and label. Justice Sotomayor cautioned that “cleverly designed surveys could also prompt such confusion by making consumers think about complex legal questions around permission that would not have arisen organically out in the world.” *Id.* at 164.

248. Parties often challenge both the admissibility of survey evidence and the weight it should be given. See Vasu & Vasu, *supra* note 239, at 71. The burden lies with the survey proponent to demonstrate that the methodology is sound (i.e., the chosen universe is appropriate and the questionnaire is adequate for the issues in dispute). *Id.* at 71–72.

249. *Id.*

contracts marked by informational and bargaining asymmetries. A rebuttable presumption framework may preserve the informational value of these signals while ensuring that the opposing party can test and rebut them with contextual evidence, reducing the risk that statistical “ordinary meaning” hardens into unjust outcomes.

In contracts between sophisticated parties, courts may treat majoritarian signals from LLMs as admissible empirical evidence supporting a party’s interpretation. When such signals strongly support one reading of the contract, courts might adopt a rebuttable presumption in its favor. The burden would then shift to the opposing party to offer countervailing evidence — such as industry usage or precontractual communications — showing that the term conveyed a different meaning.²⁵⁰

Different safeguards may be appropriate in boilerplate contracts, where structural imbalances in bargaining power and information prevail.²⁵¹ Drafters of such agreements — typically firms — are often incentivized to include one-sided or obscure terms that non-drafting parties neither expect nor understand. To counteract this, courts have developed the reasonable expectations doctrine, which permits judges to invalidate terms that a reasonable party in the recipient’s position would not have foreseen.²⁵²

250. *Supra* note 117 and accompanying text; Choi, *Measuring Clarity*, *supra* note 10, at 43 (discussing legal presumptions).

251. Standard-form contracts, commonly used in mass-market transactions, are offered by businesses to consumers on a “take-it-or-leave-it” basis, with consumers often neither reading nor understanding the terms. For critical discussion of such contracts, see generally Friedrich Kessler, *Contracts of Adhesion — Some Thoughts About Freedom of Contract*, 43 COLUM. L. REV. 629 (1943); Randy E. Barnett, *Consenting to Form Contracts*, 71 FORDHAM L. REV. 627 (2002); John D. Calamari, *Duty to Read — A Changing Concept*, 43 FORDHAM L. REV. 341 (1974); Melvin Aron Eisenberg, *The Limits of Cognition and the Limits of Contract*, 47 STAN. L. REV. 211 (1995); Robert A. Hillman & Jeffrey J. Rachlinski, *Standard Form Contracting in the Electronic Age*, 77 N.Y.U. L. REV. 429 (2002); Nathan Isaacs, *The Standardizing of Contracts*, 27 YALE L.J. 34 (1917); Russell Korobkin, *Bounded Rationality, Standard Form Contracts, and Unconscionability*, 70 U. CHI. L. REV. 1203 (2003); Arthur Allen Leff, *Unconscionability and the Code — The Emperor’s New Clause*, 115 U. PA. L. REV. 485 (1967); Michael I. Meyerson, *The Reunification of Contract Law: The Objective Theory of Consumer Form Contracts*, 47 U. MIAMI L. REV. 1263 (1993); John E. Murray, Jr., *The Parol Evidence Process and Standardized Agreements Under the Restatement (Second) of Contracts*, 123 U. PA. L. REV. 1342 (1975); John E. Murray, Jr., *The Standardized Agreement Phenomena in the Restatement (Second) of Contracts*, 67 CORNELL L. REV. 735 (1982); Todd D. Rakoff, *Contracts of Adhesion: An Essay in Reconstruction*, 96 HARV. L. REV. 1174 (1983); W. David Slawson, *The New Meaning of Contract: The Transformation of Contracts Law by Standard Forms*, 46 U. PITT. L. REV. 21 (1984); W. David Slawson, *Standard Form Contracts and Democratic Control of Lawmaking Power*, 84 HARV. L. REV. 529 (1971); James J. White, *Form Contracts Under Revised Article 2*, 75 WASH. U. L. Q. 315 (1997).

252. Courts have developed many doctrines aimed at scrutinizing and potentially invalidating unfair terms in standard-form contracts. See Ayres & Schwartz, *supra* note 117, at 556 (classifying the types of interventions into substantive and procedural); Wayne R. Barnes, *Toward a Fairer Model of Consumer Assent to Standard Form Contracts: In Defense of Restatement Subsection 211(3)*, 82 WASH. L. REV. 227, 229 (2007); Michael I. Meyerson, *The Reunification of Contract Law: The Objective Theory of Consumer Form Contracts*, 47 U.

Despite its theoretical appeal, the doctrine remains underutilized — largely due to the absence of empirical tools to assess consumer expectations.²⁵³ Without such means, courts risk applying the doctrine arbitrarily, undermining both legal predictability and market efficiency.²⁵⁴ In this context, LLM-derived majoritarian signals may provide valuable support. By reflecting dominant drafting practices in similar agreements, LLMs can help identify terms that deviate from normative baselines.

For instance, in October 2024, X amended its terms of service to impose \$15,000 in liquidated damages per platform rule violation.²⁵⁵

MIAMI L. REV. 1263, 1265 (1993). Given that standard-form contracts deviate from the traditional notion of mutual, consensual obligations, courts tend to rely on objective criteria when evaluating these agreements and “reasonable expectation” is the commonest objective benchmark for making such assessment. There are different doctrinal levers for the reasonable expectation benchmark. Originally, this benchmark emerged in the common law, usually in the context of insurance contracts. *See* Stephen J. Ware, *A Critique of the Reasonable Expectations Doctrine*, 56 U. CHI. L. REV. 1461, 1466–75 (1989); Robert A. Hillman & Jeffrey J. Rachlinski, *Standard Form Contracting in the Electronic Age*, 77 N.Y.U. L. REV. 429, 459–60 (2002). Additional grounding appears in the RESTATEMENT (SECOND) OF CONTRACTS § 211(3), which provides that “[w]here the [business] has reason to believe that the [consumer] would not [assent] if he knew that the writing contained a particular term, the term is not part of the agreement.” RESTATEMENT (SECOND) OF CONTRACTS § 211 cmt. f (AM. L. INST. 1979). The accompanying commentary to the provision further emphasizes that “[a]lthough customers typically adhere to standardized agreements and are bound by them without even appearing to know the standard terms in detail, they are not bound to unknown terms which are beyond the range of reasonable expectation.” *Id.* § 211 cmt. f (emphasis added).

253. *See, e.g.*, Barnes, *supra* note 252, at 229; Ayres & Schwartz, *supra* note 117, at 556–60.

254. Ayres & Schwartz, *supra* note 117, at 558–59. Consequently, courts have historically been hesitant to invalidate unexpected terms except in the most egregious of cases. *Id.* at 556–59. The primary vehicle court use to protect consumers from one-sided terms is not the reasonable expectation doctrine but rather the unconscionability doctrine. *See* U.C.C. § 2-302 (AM. L. INST. & UNIF. L. COMM’N. 2022); RESTATEMENT (SECOND) OF CONTRACTS § 208 (AM. L. INST. 1979). Even this doctrine is rarely used, however. *See, e.g.*, Hillman & Rachlinski, *supra* note 251, at 457–58 (2002) (“Courts generally find unconscionability when the bargaining process was deficient and the substantive terms oppressive When a form contains incomprehensible boilerplate, fine print, or otherwise hidden terms that undermine the user’s purpose of contracting or otherwise ‘shock the conscience,’ courts unhesitatingly apply unconscionability. Not surprisingly, when the context is not so stark the judicial approach is less predictable.”).

255. The complete contract provision was:

Liquidated Damages: Protecting our users’ data and our system resources is important to us. You further agree that, to the extent permitted by applicable law, if you violate the Terms, or you induce or knowingly facilitate others to do so, in addition to all other legal remedies available to us, you will be jointly and severally liable to us for liquidated damages as follows for requesting, viewing, or accessing more than 1,000,000 posts (including reply posts, video posts, image posts, and any other posts) in any 24-hour period — \$15,000 USD per 1,000,000 posts. You agree that these amounts are (i) a reasonable estimate of our damages; (ii) not a penalty; and (iii) not otherwise limiting of our ability to recover from you or others under any legal or

To assess whether this figure would be reasonably expected, we conducted a masking and autocompletion experiment using ChatGPT.²⁵⁶ We input the clause, masked the numerical value, and prompted the model to complete it. Across multiple runs, the model consistently generated \$10,000.²⁵⁷

While not a formal empirical study, this exercise revealed a plausible benchmark. If \$10,000 reflects prevailing practice, the deviation to \$15,000 — particularly if buried in fine print or disproportionately punitive — might support a presumption of unreasonableness.²⁵⁸ Courts could then shift the burden to the drafter to justify the term’s clarity, salience, and fairness.²⁵⁹

Such procedural mechanisms — burden-shifting, heightened scrutiny, or contextual evaluation — ensure that LLM-derived signals inform, but do not control, judicial reasoning. Courts retain discretion to consider party characteristics, term visibility, and substantive equity when determining whether to follow the signal.

B. Copyright Scope

Majoritarian signals in LLMs also offer potential value in refining the scope of copyright protection.²⁶⁰ Under copyright law, protection extends only to *original* expressions — those that are independently created and possess a minimal degree of creativity.²⁶¹ By contrast, standard or commonplace expressions remain in the public domain.²⁶² This distinction is doctrinally critical but notoriously difficult to apply

equitable theory or claim, including but not limited to statutory damages and/or equitable relief. You further agree that repeated violations of these Terms will irreparably harm and entitle us to injunctive and/or other equitable relief, in addition to monetary damages.

Terms of Service, X, <https://x.com/en/tos> [<https://perma.cc/XSM6-LUWD>].

256. The court’s invalidation of unreasonably large, liquidated damages is a form of semi-mandatory regulation, what Katz and Zamir called “unidirectional immutability.” Ori Katz & Eyal Zamir, *Substituting Invalid Contract Terms: Theory and Preliminary Empirical Findings*, 48 L. & SOC. INQUIRY 780, 781 (2023); see RESTATEMENT (SECOND) OF CONTRACTS § 356 (AM. L. INST. 1981).

257. ChatGPT’s full reaction was: “A typical approach for setting liquidated damages involves choosing a figure that reasonably reflects the estimated harm or loss incurred by excessive access. for a clause like this, you might see an amount such as: ‘\$10,000 USD per 1,000,000 posts’”

258. This is because the majoritarian benchmark extracted from the LLM is itself suspected in these scenarios. Cf. Dagan & Heller, *supra* note 61, at 31 (noting that “a genuinely liberal contract law, committed to empowering people’s autonomy, would be careful *not to follow* majoritarian preferences” in power-imbalanced cases); Frischmann & Vardi, *supra* note 157, at 24 (“[W]hat is reasonable, unreasonable and egregiously unreasonable is a moving target, shaped and normalized by common practices.”).

259. This presumption extends from the traditional *contra proferentem* principle. See *supra* note 208.

260. See generally Hacoheh & Elkin-Koren, *supra* note 10.

261. *Feist Publ’ns, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340, 345 (1991).

262. See generally Hacoheh & Elkin-Koren, *supra* note 10.

in practice, often resulting in the overprotection of unoriginal or generic content.²⁶³

Foundation models present a novel opportunity to improve the consistency of originality determinations. Because these models are trained on vast corpora of textual, visual, and auditory works, they implicitly encode dominant expressive patterns across a wide range of genres and domains. By revealing those majoritarian patterns, LLMs can assist courts in identifying when a particular expression is sufficiently conventional, expected, or widely used so that it should remain outside the scope of copyright protection.²⁶⁴

1. Normative Framework

Copyright law grants exclusive rights to authors of expressive works.²⁶⁵ However, to fulfill its constitutional mandate to “promote the Progress of Science and useful Arts,”²⁶⁶ the scope of protection must be limited.²⁶⁷ Such limitations are essential because creative progress depends on cumulative innovation and sustained engagement with existing cultural materials.²⁶⁸

263. See, e.g., James Gibson, *Risk Aversion and Rights Accretion in Intellectual Property Law*, 116 YALE L.J. 882, 884–85 (2007) (arguing how the unpredictability of liability leads risk averse users to seek a license where none is needed); see also NEIL WEINSTOCK NETANEL, COPYRIGHT’S PARADOX 6–7 (2008) (discussing how copyright’s exponential growth is in “ungainly distension” with free speech); Jessica Litman, *Billowing White Goo*, 31 COLUM. J.L. & ARTS 587, 587 (2008) (noting how fair use has failed to keep up with huge expansions in copyright protections); John Tehranian, *Infringement Nation: Copyright Reform and the Law/Norm Gap*, 2007 UTAH L. REV. 537, 543–48 (explaining how ordinary behaviors, such as unauthorized video recording, could be criminal infringement under current copyright law); Yochai Benkler, *Free as the Air to Common Use: First Amendment Constraints on Enclosure of the Public Domain*, 74 N.Y.U. L. REV. 354, 357–58 (1999) (weighing the public interest costs of broadening copyright law over information production and exchange processes); Gideon Parchomovsky & Alex Stein, *Originality*, 95 VA. L. REV. 1505, 1513 (2009).

264. See generally Hacoen & Elkin-Koren, *supra* note 10.

265. *Mazer v. Stein*, 347 U.S. 20, 29 (1954).

266. U.S. CONST. art. I, § 8, cl. 8.

267. Anna Robilant & Talha Syed, *Property’s Building Blocks: Hohfeld in Europe and Beyond*, in THE LEGACY OF WESLEY HOHFELD: EDITED MAJOR WORKS, SELECT PERSONAL PAPERS, AND ORIGINAL COMMENTARIES 229 (Henry Smith et al. eds., 2022). Julie E. Cohen, *Lochner in Cyberspace: The New Economic Orthodoxy of “Rights Management,”* 97 MICH. L. REV. 462, 498–502 (1998); Mark A. Lemley, *Property, Intellectual Property, and Free Riding*, 83 TEX. L. REV. 1031, 1037–38 (2005).

268. This dynamic applies to all intellectual property regimes such as trademark, copyright, and patent laws. Because information goods are non-rival and can benefit society at large (owners and non-owners alike), intellectual property laws must carefully balance exclusion with access. Like copyright originality and limitation doctrines, many other intellectual property doctrines achieve this balance by deferring to majoritarian benchmarks. Patent law, for example, affords legal protection to innovations based on how far they deviate from the majoritarian industry practices and “mindsets” when the patent application is filled. See generally Laura G. Pedraza-Fariña & Ryan Whalen, *A Network Theory of Patentability*, 87 U. CHI.

Accordingly, copyright protects only the *original* aspects of expressive works.²⁶⁹ To qualify as original, a work must (1) originate with the author and (2) exhibit a minimal degree of creativity.²⁷⁰ Facts, for example, are not protectable, as they are not the product of authorship. Courts interpret this creativity requirement broadly but still exclude standardized, conventional, or abstract elements from protection.

Judicial application of this threshold often turns on deference to *generic expressive patterns*. In *Feist Publications, Inc. v. Rural Telephone Service Co.*,²⁷¹ the Supreme Court held that an alphabetical listing of telephone subscribers lacked originality because it followed a “time-honored tradition” and did not reflect creative choice.²⁷² Similarly, abstract or low-detail patterns are excluded as unprotectable “ideas,” not expressive implementations.²⁷³

Originality also plays a central role in copyright enforcement. Under the analytical framework established in *Computer Associates International, Inc. v. Altai, Inc.*,²⁷⁴ courts adjudicating infringement claims first “filter out” unprotectable elements — such as factual content, ideas, or *scènes à faire* — before comparing the remaining expressive

L. REV. 63 (2020). To assess non-obviousness in patent law; courts need to apprise whether an alleged invention was obvious in the eyes of the person having ordinary skill in the art (“PHOSITA”) when the patent application was filled. See John O. Tresansky, *PHOSITA — The Ubiquitous and Enigmatic Person in Patent Law*, 73 J. PAT. & TRADEMARK OFF. SOC’Y 37, 37 (1991). See generally Joseph P. Meara, *Just Who Is the Person Having Ordinary Skill in the Art? Patent Law’s Mysterious Personage*, 77 WASH. L. REV. 267 (2002). The PHOSITA standard is notoriously tricky to apply, however, mainly because human factfinders are susceptible to hindsight bias, namely the tendency to consider non-obvious advances trivial in retrospect. Cf. Baruch Fischhoff, *For Those Condemned To Study the Past: Heuristics and Biases in Hindsight*, in JUDGMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES 335, 341 (Daniel Kahneman, Paul Slovic, & Amos Tversky, eds., 1982); ROBERT HARMON, HARMON ON PATENTS: BLACK LETTER LAW & COMMENTARY 466 (2007). In the famous English *Dyson v. Hoover* case, for example, the court considered Dyson’s decision to use cyclonic separation for purifying dirty air in vacuum cleaners innovative because it was against the industry standard in this time, which overwhelmingly favored filtering bags. See *Dyson v. Hoover* [2001] EWCA (Civ) 1440; PETER D. ROSENBERG, PATENT LAW FUNDAMENTALS § 9.02[2][a][i] (2000). Trademark law also defers to majoritarian forms of commercial expression. Thus, when a substantial portion of consumers use a trademark to signify a general product or service category rather than a specific brand, it loses its legal protection. See generally *supra* note 82. Majoritarian bias signals in foundational models can, in principle, also inform trademark and patent law scope delineating doctrines. However, we will not consider them here because of methodological limitations.

269. 17 U.S.C. § 102(a); *Feist Publ’ns, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340, 345 (1991); *Harper & Row, Publishers, Inc. v. Nation Enters.*, 471 U.S. 539, 547 (1985).

270. See *Feist*, 499 U.S. at 345, 359 (1991); Hacothen & Elkin-Koren, *supra* note 10, at 573–74.

271. 499 U.S. at 363 (1991).

272. *Id.* at 363. (non-creative expression is an “age-old practice, firmly rooted in tradition and so commonplace that it has come to be expected as a matter of course”).

273. *Id.* at 345; PAUL GOLDSTEIN, GOLDSTEIN ON COPYRIGHT § 2.3 (3d ed. 2005).

274. 982 F.2d 693 (2d Cir. 1992).

material with the allegedly infringing work.²⁷⁵ The more original the plaintiff's expression, the more robust its legal protection.²⁷⁶

During this *filtration stage*, courts apply various limitation doctrines to distinguish between protectable and unprotectable content.²⁷⁷ In *Acuff-Rose Music, Inc. v. Jostens, Inc.*,²⁷⁸ the court excluded the phrase "You've Got to Stand for Something" on the grounds that it was overly clichéd. In *Lotus Development Corp. v. Borland International, Inc.*,²⁷⁹ the First Circuit found that a command menu hierarchy lacked protection because widespread use had transformed it into an unprotectable "method of operation."²⁸⁰

Courts also considered genericity at the *comparison stage* of *Altai*, assessing whether two works were "substantially similar."²⁸¹ The more generic the expressive elements of the plaintiff's work, the higher the threshold for establishing similarity.²⁸² In *Apple Computer, Inc. v. Microsoft Corp.*,²⁸³ the court found that Apple's graphical interfaces were widely used to the point of becoming an industry standard, which rendered them unprotected.²⁸⁴ Similarly, in *Oracle America, Inc. v. Google, Inc.*, the Court deemed Google's verbatim copying of certain code permissible, in part because the copied lines were already in widespread use.²⁸⁵

Moreover, originality is relevant in fair use analysis.²⁸⁶ Derivative works that display significant creative transformation are more likely

275. Mark A. Lemley, *Our Bizarre System for Proving Copyright Infringement*, 57 J. COPYRIGHT SOC'Y 719, 748 (2010).

276. Mark A. Lemley, *Convergence in the Law of Software Copyright*, 10 HIGH TECH. L. J. 1, 28 (1995); *Apple Computer v. Microsoft Corp.*, 35 F.3d 1435, 1445 (9th Cir. 1994).

277. *SAS Institute Inc. v. World Programming Ltd.*, 64 F. Supp. 3d 755, 775–76 (E.D.N.C. 2014); Pamela Samuelson, *Reconceptualizing Copyright's Merger Doctrine*, 63 J. COPYRIGHT SOC'Y 417 (2016); Jessica D. Litman, *The Public Domain*, 39 EMORY L. J. 965, 975 (1990).

278. 155 F.3d 140, 143 (2d Cir. 1998).

279. 49 F.3d 807, 815–18 (1st Cir. 1995).

280. Not all the judges prescribed to the dynamic approach discussed in Judge Boudin's concurrence. *Id.* at 819–22. The majority believed that that Lotus 1-2-3's command menu hierarchy didn't become unprotected due to overuse but was unprotected from the start.

281. The courts analysis of "substantial similarity," and "fair use" requires them to implicitly consider the originality of the allegedly infringed copyrighted works. See Lemley, *supra* note 276, at 28; *Apple Computer*, 35 F.3d at 1443 (9th Cir. 1994); 17 U.S.C. § 107(1)–(2).

282. Lemley, *supra* note 275, at 748.

283. 35 F.3d at 1438, 1443.

284. The originality problem there was due to simplicity, not genericity, but the same logic should apply to all non-original aspects of copyrighted works no matter how they came to be so.

285. *Oracle America, Inc. v. Google, Inc.*, 593 U.S. 1, 24 (2021); see also Peter S. Menell, *An Analysis of the Scope of Copyright Protection for Application Programs*, 41 STAN. L. REV. 1045, 1066–67, 1101 (1989). The court found the use to be permissible under fair use analysis, but prominent scholars have long argued that such generic content — at least in the dominion of computer code — should be excluded *ab initio* from copyright protection.

286. Mark A. Lemley, *The Economics of Improvements in Intellectual Property Law*, 75 TEX. L. REV. 989, 1077 (1997); Lemley, *supra* note 276, at 30.

to be deemed fair use.²⁸⁷ Thus, courts must not only evaluate the originality of the plaintiff's work but also consider whether the defendant's use adds sufficient new expression or meaning.

From a normative perspective, then, majoritarian signals from LLMs are well-suited to assist courts in navigating the boundaries of copyright protection.²⁸⁸ However, the nature of their role differs significantly from that in contract interpretation. In contract law, majoritarian signals are primarily lexical, helping courts infer the ordinary meaning of words used in agreements, especially when shared subjective intent is unavailable.²⁸⁹ In copyright law, by contrast, the normative function of majoritarian signals is ontological: they help courts assess whether specific expressive elements are *commonplace* in a given domain. Rather than merely clarifying word meaning, LLMs can help determine whether a given phrase, image, or stylistic device is so generic that it fails to meet the originality threshold. This task — identifying and quantifying expressive commonality — was historically impossible to perform systematically but is now increasingly feasible with the tools provided by foundation models.

By uncovering dominant expressive patterns across massive cultural corpora, LLMs may help courts identify unprotectable content with greater consistency and precision. Such signals could support courts in filtering out *scènes à faire* and other generic elements, thereby reinforcing the protection of truly original contributions. In turn, this use of majoritarian benchmarks could improve both doctrinal coherence and normative fairness by helping to ensure that copyright protection reflects genuine creative efforts.²⁹⁰

Looking forward, there may be grounds for rethinking copyright's doctrinal architecture to better accommodate these new empirical tools. Policymakers might even consider amending interpretive standards to explicitly allow courts to rely on LLM-generated majoritarian signals when assessing originality, filtration, or substantial similarity. The viability of such reforms, however, will depend on the methodological and procedural challenges discussed in the following sections.

287. Pierre N. Leval, *Toward a Fair Use Standard*, 103 HARV. L. REV. 1105, 1107–11 (1990); *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 578–79 (1994).

288. See generally Hacohen & Elkin-Koren, *supra* note 10 (discussing the benefits of GenAI use in copyright law); Michal Shur-Ofry, *Popularity as a Factor in Copyright Law*, 59 U. TORONTO L. J. 525 (2009) (advocating for incorporating indicia of popularity/majority opinion into copyright analysis).

289. See *supra* Section III.A.1.

290. See generally Hacohen & Elkin-Koren, *supra* note 10; Dan L. Burk, *Cheap Creativity and What It Will Do*, 57 GA. L. REV. 1669 (2023).

2. Methodology

Courts have traditionally assessed copyrightability benchmarks intuitively rather than empirically, resulting in vagueness, inconsistency, and overprotection.²⁹¹ Foundation models now offer an opportunity to surface dominant expressive patterns across cultural corpora, providing an empirical basis for distinguishing between generic and original content.

a. Inpainting and Visual Genericity

Inpainting — the visual analogue of autocompletion — is one promising approach for empirically distinguishing between generic and original content²⁹² Using foundation models trained on vast image datasets (e.g., Stable Diffusion), courts or litigants could assess whether particular visual elements are statistically prevalent — or, instead, anomalous — within a given genre or corpus.

To illustrate this approach, we conducted two inpainting experiments using Stable Diffusion, both depicted in Figure 3. In the first, we removed a dog from a photo of a bench and asked the model to reconstruct the image. The model filled in the empty space with a vacant bench, suggesting that the model had encountered more images of benches without dogs than with them and thus defaulted to the dominant pattern.

In the second experiment, we masked the apple from René Magritte’s *The Son of Man*²⁹³ and prompted the model to reconstruct the missing region. The model generated a generic male face in place of the apple — again, defaulting to the statistically dominant pattern in its training data.²⁹⁴ In both cases, the reconstructed content revealed what the model “expected” to see given prevailing visual conventions.

291. Michael J. Meurer & Peter Menell, *Notice Failure and Notice Externalities*, 5 J. LEGAL ANALYSIS 1, 5 (2013).

292. Rombach et al., *supra* note 33, at 8 (“[I]npainting is the task of filling masked regions of an image with new content either because parts of the image are corrupted or to replace existing but undesired content within the image.”).

293. Cath Pound, *Magritte and the Subversive Power of His Pipe*, BBC CULTURE (Dec. 5, 2017), <https://www.bbc.com/culture/article/20171205-magritte-and-the-subversive-power-of-his-pipe> [<https://perma.cc/5TB2-QCBM>].

294. We tried this exercise also with adding to the model the textual prompt “Magritte The Son of Man,” but the outputs were still dominated by men’s faces. This outcome may also demonstrate that the model’s training data might be biased in favor of images of men wearing a suit and a hat, rather than images of women, but it is difficult to tell. Given that the body shape and other cues in the outfit are clearly masculine, the model may properly recognize the appropriate gender.

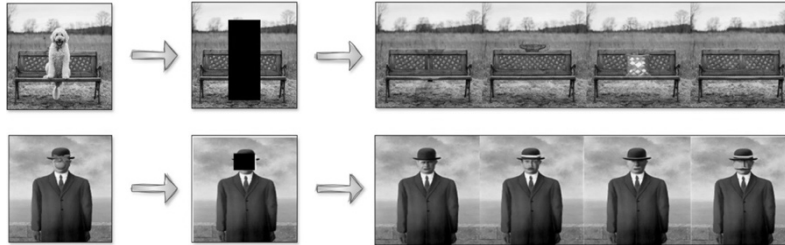


Figure 3: An inpainting exercise to reconstruct the bench from a photo of a dog on a bench (Top) and an apple in René Magritte’s painting *The Son of Man* (Bottom).

This methodology could assist courts in applying doctrines such as *scènes à faire*.²⁹⁵ For example, in *Steinberg v. Columbia Pictures*,²⁹⁶ the court examined whether a movie poster for *Moscow on the Hudson* impermissibly copied original elements from Saul Steinberg’s famous *New Yorker* cover, “View of the World from 9th Avenue.” The court noted that features such as “buildings, pedestrians, vehicles, lampposts, and water towers” may be *scènes à faire* — ubiquitous in depictions of New York City — and thus may be unprotected, though specific depictions of these commonplace features could still be protected.²⁹⁷ It further found that other details, such as “shadows and streetlights,” were likely original and thus protectable.²⁹⁸

We replicated this analysis using Stable Diffusion, shown below in Figure 4. After masking Steinberg’s shadows and streetlights, we prompted the model to reconstruct them. Consistent with the court’s reasoning, the model defaulted to generic urban imagery — reproducing buildings but not Steinberg’s distinctive lighting details.

295. In U.S. copyright law, the *scènes à faire* doctrine holds that copyright does not protect elements of a work that are standard, stock, or indispensable to a particular genre, setting, or subject matter. *SAS Institute Inc. v. World Programming Ltd.*, 64 F.4th 1319, 1325–26 (Fed. Cir. 2023) (“[C]opyright protection extends only to the expression of an idea, not to the underlying idea itself. . . . Additionally, other doctrines of copyright law detail what elements are not protectable, including *scènes à faire* elements, material in the public domain, factual material, and elements under the merger doctrine.”). Under the *scènes à faire* doctrine, expressions indispensable and naturally associated with the treatment of a given idea “are treated like ideas and are therefore not protected by copyright.” *Rice v. Fox Broad. Co.*, 330 F.3d 1170, 1175 (9th Cir. 2003); see *Gaiman v. McFarlane*, 360 F.3d 644, 659 (7th Cir. 2004); *Litman*, *supra* note 277, at 996–67 (noting that “*scènes à faire* seems to lie more in their triteness than their necessity”).

296. 663 F. Supp. 706 (S.D.N.Y. 1987).

297. *Id.* at 713. Clearly, buildings, pedestrians, vehicles, lampposts, and water towers can also be portrayed in an original form. But the small, low-detail panoramic angle in which the city skyline portrayal was done was possibly non-original.

298. *Id.* (describing them as “errors and anomalies”).



Figure 4: Inpainting exercise on View of the World from 9th Avenue

These results are particularly compelling because the model had likely encountered Steinberg’s work in its training data.²⁹⁹ The fact that it failed to reproduce those masked elements — despite being prompted with Steinberg’s own composition — suggests that those features are *not* statistically dominant. Thus, the model’s “surprise” reinforces the legal conclusion that those elements were not generic.

b. Autocompletion and Textual Genericity

Similar techniques can be applied to textual works through auto-completion. By masking or partially obscuring a phrase and prompting a language model to complete it, litigants or courts can assess the statistical likelihood of particular phrasings and thereby gauge their conventionality.

For example, a prior study found that LLMs predicted the endings of different idioms with varying levels of success.³⁰⁰ These results suggest that some idioms are more widely used than others. Copyright law generally excludes such familiar expressions from protection — as exemplified in the *Acuff-Rose Music v. Jostens* determination that the phrase “You’ve Got to Stand for Something” was too clichéd to be original.³⁰¹ Although copyright doctrine does not currently impose numeric thresholds for genericity, probabilistic measures of linguistic conventionality could add nuance to judicial assessment of whether a particular expression is too commonplace to merit protection.

Autocompletion can thus operationalize the intuition that certain phrasings are “commonplace.” Phrases with high completion probabilities may serve as proxies for majoritarian usage, signaling

299. As was implicitly noted in the autocompletion exercises mentioned above, majoritarian signals are especially instructive when they “go against” the features that are the subject of the judicial scrutiny. *See supra* note 258 and accompanying text.

300. Haviv, *supra* note 31, at 254–55.

301. 155 F.3d 140, 143 (2d Cir. 1998); *see* 37 C.F.R. § 202.1(a) (2004); *Perma Greetings, Inc. v. Russ Berrie & Co.*, 598 F. Supp. 445, 448–49 (E.D. Mo. 1984).

unprotectable content under the merger doctrine, *scènes à faire*, or the originality threshold itself.

c. Toward Robust Methodologies

The extraction of majoritarian signals from LLMs for copyright purposes remains in its infancy. These methodologies, which blend textual and visual analysis, could eventually guide not only courts but also the Copyright Office in more consistently assessing the scope of protection.³⁰²

Over time, refined empirical benchmarks derived from foundation models could help operationalize longstanding doctrinal goals — ensuring that copyright protection remains limited to truly original contributions, while preserving robust access to expressive materials that form the foundation of cultural progress.

3. Legal Procedure

In the context of copyright infringement litigation, courts should adopt procedural safeguards analogous to those discussed in Section III.A.3 *supra*. For example, litigants might submit inpainting exercises as empirical evidence to support *scènes à faire* arguments. In such cases, courts could apply familiar procedural mechanisms — including legal presumptions, burden-shifting, and opportunities for cross-examination — to ensure fair and transparent evaluation.

However, as discussed in Section III.B.1. *supra*, the normative function of majoritarian signals in copyright differs from their role in contract law. In contract interpretation, majoritarian signals help judges ascertain the ordinary meaning of disputed terms, much like dictionaries or corpus linguistics, which courts often consult with limited evidentiary scrutiny. In copyright cases, by contrast, majoritarian signals

302. Hacoen & Elkin-Koren, *supra* note 10, at 556. Note that the ability of majoritarian bias signals to apprise copyright originality may also be valuable to artists and consumers and facilitate markets for copyrighted works. Just as the assessment of genericity in creative expressions may assist courts or the copyright office in determining the copyrightability of expressive elements, this metric may also prove beneficial for artists or merchants dealing with creative works. Creators may want to gauge the “originality” of their expressive works, either out of intellectual or artistic curiosity or to obtain an objective appraisal. With such measurements, for example, it is not difficult to envision companies seeking and generously compensating more original works of human authorship that may have a higher market value. It is currently unclear whether LLM training requires the copyright owners’ authorization or whether it is lawful as it constitutes non-expressive fair use. See Mark A. Lemley & Bryan Casey, *Fair Learning*, 99 TEX. L. REV. 743, 744–46 (2021). But even if training GenAI’s systems on expressive input is lawful, companies may still wish to encourage “original” creation instead of generic AI-assisted creation. Original creations will give GenAI system developer a competitive advantage by allowing them to continuously improve their models. The better and more widely used GenAI systems will be, the more precious the value of original human creation will become.

serve an ontological function: they help determine whether expressive elements are so commonplace as to fall outside protection, a judgment that may warrant stricter safeguards for reliability and admissibility.³⁰³

This concern is analogous to the longstanding evidentiary standards applied to consumer surveys in trademark litigation.³⁰⁴ Like majoritarian signals, surveys offer statistical insights into public perception.³⁰⁵ To be admissible, they must meet rigorous criteria, such as proper sampling, clear question design, and methodological reliability.³⁰⁶ Courts could require similar standards for majoritarian signal methodologies, particularly in assessing expressive genericity. Ideally, the reliability of such signals would be validated through independent testing, cross-model benchmarking, or other scientific controls.

The same procedural concerns extend to the U.S. Copyright Office, especially as it assumes a gatekeeping role in determining copyrightability. Works eligible for copyright protection can also be registered in the Copyright Office. Registration is not required for copyright eligibility but is necessary for filing copyright infringement lawsuits.³⁰⁷

In principle, the Office could issue formal guidelines specifying the permissible use of majoritarian signal methodologies in evaluating originality. Such guidelines would need to articulate acceptable methodological frameworks, transparency requirements, and mechanisms for authors or experts to audit, challenge, or appeal determinations made using these tools.³⁰⁸

Although the Copyright Office recently updated its guidance on the copyrightability of AI-generated works, it has not yet addressed the use of foundation models as *evaluative instruments* — even though such use could be particularly appropriate in that context.³⁰⁹ Under the Office’s current approach, AI-generated works that lack sufficient human authorship are categorically excluded from protection.³¹⁰ Majoritarian signals could help the registrar flag works as generic and deny registration — or, at a minimum, treat the work’s generic elements as

303. Choi, *Measuring Clarity*, *supra* note 10, at 15.

304. *See id.* at 51 (discussing the limitations of surveys).

305. *Id.*

306. *See* Vasii & Vasii, *supra* note 239, at 71; *see also* L. Rierson & Robert H. Thornburg, *Trademark Survey Evidence: Review of Current Trends in the Ninth Circuit*, 21 SANTA CLARA HIGH TECH. L.J. 715, 743 (2005); Shari S. Diamond & David J. Franklyn, *Trademark Surveys: An Undulating Path*, 92 TEX. L. REV. 2029, 2037 (2014).

307. *See generally* Zvi S. Rosen, *Examining Copyright*, 69 J. COPYRIGHT SOC’Y USA 481 (2024).

308. The Copyright Office could incorporate more stringent guidelines, as the Patent Office does in its detailed administrative review process. *See* Michael J. Flibbert & Sarah E. Craven, *Appellate Review of the AIA’s New Patent-Challenge Proceedings*, FINANCIER WORLDWIDE (Mar. 2013), <https://www.financierworldwide.com/appellate-review-of-the-aias-new-patent-challenge-proceedings> [<https://perma.cc/H9N2-J5B8>].

309. *See generally* U.S. COPYRIGHT OFF., COPYRIGHT AND ARTIFICIAL INTELLIGENCE, PART 2: COPYRIGHTABILITY (2025).

310. *Id.* at 7.

unprotectable. Thus, they could aid the Copyright Office bypass the need to rely on assessing creativity in prompt engineering and instead compare the originality of the work to the preexisting materials from which the model has already learned.

Despite this theoretical promise, procedural standardization is unlikely to emerge in the near term. Several challenges stand in the way. First, the underlying technologies are evolving rapidly, with LLM architectures, training datasets, and alignment methods changing frequently. Any fixed procedural benchmark risks obsolescence: a rule defined today may fail to reflect tomorrow's capabilities.³¹¹ Second, and perhaps most crucially, there is the problem of compliance circularity. Once a procedural standard is articulated, there is a risk that litigants and technologists will tailor methodologies merely to *satisfy* the standard, rather than to improve reliability. This dynamic could produce a veneer of methodological rigor without genuine epistemic improvement.³¹²

In practice, then, the reliability of majoritarian signals in copyright litigation will likely be determined on a case-by-case basis, with courts evaluating admissibility, weight, and probative value in light of context-specific factors. Over time, a common law of evidentiary standards may emerge, shaped by precedent, professional norms, and the iterative refinement of computational tools.

IV. CONCLUSION

GenAI foundation models have ushered in a transformative moment — not only for technology and society, but also for the law. Traditionally, legal analysis has struggled to operationalize vague or open-ended standards that depend on what is “reasonable,” “ordinary,” or “widely understood.” Courts often rely on heuristics — surveys, dictionaries, expert witnesses — to approximate majoritarian understanding, but these tools are frequently inconsistent, expensive, or poorly aligned with how language and norms evolve. GenAI, with its ability to extract statistical regularities from vast corpora of real-world data and reveal how ordinary people speak, behave, and expect others to behave, offers a new, empirically grounded alternative.

As discussed above, contrary to early skepticism in legal scholarship, majoritarian signals in GenAI are not merely technical flaws to be mitigated but epistemically valuable indicators of collective human judgment. When approached critically and carefully, these signals can offer courts a scalable method for interpreting legal standards that

311. See Paul W. Grimm, Maura R. Grossman & Gordon V. Cormack, *Artificial Intelligence as Evidence*, 19 NW. J. TECH. & INTELL. PROP. 9, 82–83 (2021).

312. *Id.*

already depend on majoritarian reference points — from “ordinary meaning” in contracts to “generic expression” in copyright law. GenAI models, in short, do not create majoritarian reasoning in law; what they do is make it measurable.

However, the use of these tools is not without risks. Their outputs can reflect — and sometimes amplify — deep social asymmetries. Their reliability can be undermined by poor prompt design, data skew, or architectural idiosyncrasies. And their evidentiary role in court raises complex procedural questions about admissibility, expertise, and transparency. If GenAI signals are to play a legitimate role in adjudication, they must be subjected to the same kind of methodological discipline and procedural scrutiny as other forms of scientific or statistical evidence.

This Article proposed a framework for doing just that. First, it articulated the normative rationale for considering majoritarian signals not as conclusive answers, but as informative proxies that can support (and occasionally challenge) judicial reasoning. Second, it examined methodological best practices — including cosine similarity, autocompletion, and inpainting — that enable courts and litigants to responsibly extract majoritarian signals from foundation models. Third, it explored procedural safeguards that can ensure fairness, transparency, and accountability in the use of these tools — especially in asymmetrical contracting contexts and copyright disputes involving contested originality.

Referring to two case studies — contract interpretation and copyright scope — this Article showed that GenAI-derived signals can help courts resolve ambiguity, surface dominant cultural conventions, and ground adjudication in empirically observable practices. At the same time, it emphasized that GenAI should never be treated as a black box authority. Rather, it should function as a tool of legal interpretation: one that informs but does not displace normative judgment.

As foundation models continue to evolve, their majoritarian outputs will only become more sophisticated, customizable, and accessible. This both creates opportunities and imposes new responsibilities on the legal system. Courts, scholars, and policymakers must begin to articulate standards for when and how these tools can be used — not just to detect bias or automate tasks, but to enrich legal reasoning itself.

GenAI models should not decide legal questions, but their data-rich representations of shared meaning can help judges and litigants reason more transparently about what is ordinary, expected, or reasonable in law. If we treat these models not as threats to legal reasoning but as sources of insight — tempered by doctrine, institutional safeguards, and critical judgment — we may find in them not just problems to solve, but new ways of seeing law itself.