

Michael J. Bommarito II

Technology | Law | Finance

michael.bommarito@gmail.com | michaelbommarito.com | github.com/mjbmommar | linkedin.com/in/bommarito | [Scholar](#)

61 Publications

4,291 Citations

22 *h*-index

26 i10-index

Google Scholar, June 2026

Researcher and serial entrepreneur working at the intersection of technology, law, and finance for over 25 years. Work spans artificial intelligence, legal informatics, computational law, and quantitative finance, with research published in venues including *Science*, *Quantitative Finance*, and Cambridge University Press.

Professional Experience

President , ALEA Institute	Present
Chief Executive Officer , 273 Ventures	Present
Chief Executive Officer , Bommarito Consulting, LLC	Present
Chief Technology Officer , licens.io	Past
Chief Executive Officer & Co-Founder , LexPredict (acquired by Elevate, 2018)	2013–2018

Academic Appointments

Affiliate , CodeX , Stanford Center for Legal Informatics, Stanford University	2024–present
Fellow , CodeX , Stanford Center for Legal Informatics, Stanford University	2014–2023
Head of Research , Law Lab , Chicago-Kent College of Law, Illinois Institute of Technology	2017–present
Adjunct Professor , College of Law, Michigan State University	2012–2017
Lecturer , Center for the Study of Complex Systems, University of Michigan	2014–2016

Education

M.S.E., Financial Engineering, University of Michigan, Ann Arbor
M.A., Political Science, University of Michigan, Ann Arbor
B.S., Mathematics, University of Michigan, Ann Arbor
ABD Ph.D., Political Science, University of Michigan, Ann Arbor

Books & Edited Volumes

[1] Katz, D. M., Dolin, R., & **Bommarito, M. J.** (2021). *Legal informatics*. Cambridge University Press. ISBN: 978-1107142725 ↗

- [2] **Bommarito, M. J.** (2026). *This Is Server Country: AI, Power, and the Remaking of Rural America*. Self-published. ISBN: 979-8-9943457-3-3 (paperback), 979-8-9943457-2-6 (ebook) ↗
- [3] **Bommarito, M. J.** (2026). *The Math Inside the Machine: How Intelligence Emerges from Eleven Simple Operations*. Self-published. ISBN: 979-8-9943457-8-8 (paperback), 979-8-9943457-9-5 (ebook) ↗
- [4] **Bommarito, M. J., II** (2026). *Rough Consensus: How Engineers, Hackers, and Spies Built the Internet*. Self-published. ISBN: 979-8-9943457-7-1 (paperback), 979-8-9943457-6-4 (ebook)

Books in Preparation

- [1] **Bommarito, M. J., II** (2026). *The Commons Captured: A Critical History of Wikipedia*. Self-published. *Draft manuscript*.
- [2] **Bommarito, M. J.** (2026). *From Clay to Code: How Legal Technology Shaped the Modern World*. Self-published. *Draft manuscript*.
- [3] **Bommarito, M. J.** (2026). *From C to Windows Kernel Defense: A Practitioner's Guide to Windows Kernel Security*. Self-published. *Draft manuscript*.
- [4] **Bommarito, M. J.** (2026). *From C to Kernel Defense: A Practitioner's Guide to Linux Kernel Security*. Self-published. *Draft manuscript*.

Articles, Proceedings & Chapters

- [1] **Bommarito, M. J.**, Bommarito, J., & Katz, D. M. (2025). *What is an Agent? A Conceptual Primer and History of Agents and Agentic AI*. SSRN. ↗
- [2] **Bommarito, M. J.**, Katz, D. M., & Bommarito, J. (2025). *How to Design an AI Agent: Architectures, Protocols, and Technical Evaluation of Agentic AI Systems for Law & Finance*. SSRN. ↗
- [3] Bommarito, J., Katz, D. M., & **Bommarito, M. J.** (2025). *Governing AI Agents: Risk, Compliance, and Accountability in Law and Finance*. SSRN. ↗
- [4] Katz, D. M., **Bommarito, M. J.**, Gao, S., & Arredondo, P. (2024). *GPT-4 passes the bar exam*. *Philosophical Transactions of the Royal Society A*. Volume 382, Issue 2270, Article 20230254 ↗
- [5] Chalkidis, I., Jana, A., Hartung, D., **Bommarito, M.**, Androutsopoulos, I., Katz, D. M., & Aletras, N. (2022). *LexGLUE: A benchmark dataset for legal language understanding in English*. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (ACL 2022)*. Volume 1: Long Papers, Pages 4310-4330 ↗
- [6] **Bommarito II, M. J.** (2021). *Preprocessing Data*. *Legal Informatics (Chapter 2.3)*. In *Legal Informatics*, Katz, D. M., Dolin, R., & Bommarito, M. J. (eds.). Cambridge University Press. ↗
- [7] Coupette, C., Beckedorf, J., Hartung, D., **Bommarito, M.**, & Katz, D. M. (2021). *Measuring law over time: A network analytical framework with an application to statutes and regulations in the United States and Germany*. *Frontiers in Physics*. Volume 9, Article 658463 ↗
- [8] **Bommarito II, M. J.**, Katz, D. M., & Detterman, E. M. (2021). *LexNLP: Natural language processing and information extraction for legal and regulatory texts*. *Research Handbook on Big Data Law*. Pages 216-227, Edward Elgar Publishing ↗

- [9] Lee, E. D., Katz, D. M., **Bommarito, M. J.**, & Ginsparg, P. H. (2020). *Sensitivity of collective outcomes identifies pivotal components*. Journal of The Royal Society Interface. Volume 17, Issue 167, 20190873 [↗](#)
- [10] **Bommarito II, M. J.**, & Duran, A. (2018). *Spectral analysis of time-dependent market-adjusted return correlation matrix*. Physica A: Statistical Mechanics and its Applications. Volume 503, Pages 273-282 [↗](#)
- [11] **Bommarito II, M. J.**, & Katz, D. M. (2017). *Measuring and modeling the US regulatory ecosystem*. Journal of Statistical Physics. Volume 168, Issue 5, Pages 1125-1135 [↗](#)
- [12] Ruhl, J.B., Katz, D. M., & **Bommarito, M. J.** (2017). *Harnessing legal complexity*. Science. Volume 355, Issue 6332, Pages 1377-1378 [↗](#)
- [13] Katz, D. M., **Bommarito, M. J.**, & Blackman, J. (2017). *A general approach for predicting the behavior of the Supreme Court of the United States*. PLoS One. Volume 12, Issue 4, e0174698 [↗](#)
- [14] Candeub, A., & **Bommarito, M. J.** (2015). *Understanding the Federal Communication Commission's Policy-Making Using Big Data*. TPRC 43: The 43rd Research Conference on Communication, Information and Internet Policy. [↗](#)
- [15] Katz, D. M., & **Bommarito, M. J.** (2014). *Measuring the complexity of the law: the United States Code*. Artificial Intelligence and Law. Volume 22, Issue 4, Pages 337-374 [↗](#)
- [16] Agrawal, A., Brown, D. G., Rao, G., Riolo, R., Robinson, D. T., & **Bommarito II, M.** (2013). *Interactions between organizations and networks in common-pool resource governance*. Environmental Science & Policy. Volume 25, Pages 138-146 [↗](#)
- [17] Katz, D. M., Gubler, J. R., Zelner, J., & **Bommarito, M. J.** (2011). *Reproduction of hierarchy? A social network analysis of the American law professoriate*. Journal of Legal Education. Volume 61, Page 76 [↗](#)
- [18] Katz, D. M., **Bommarito, M. J.**, Seaman, J., Candeub, A., & Agichtein, E. (2011). *Legal n-grams? A simple approach to track the 'evolution' of legal language*. Proceedings of JURIX 2011. The 24th International Conference on Legal Knowledge and Information Systems, Vienna [↗](#)
- [19] Duran, A., & **Bommarito, M. J.** (2011). *A profitable trading and risk management strategy despite transaction costs*. Quantitative Finance. Volume 11, Issue 6, Pages 829-848 [↗](#)
- [20] **Bommarito II, M. J.**, Katz, D. M., & Zelner, J. L. (2010). *On the stability of community detection algorithms on longitudinal citation data*. Procedia-Social and Behavioral Sciences. Volume 4, Pages 26-37 [↗](#)
- [21] **Bommarito II, M. J.**, Katz, D. M., Zelner, J. L., & Fowler, J. H. (2010). *Distance measures for dynamic citation networks*. Physica A: Statistical Mechanics and its Applications. Volume 389, Issue 19, Pages 4201-4208 [↗](#)
- [22] **Bommarito, M. J.** (2010). *An Empirical Survey of the Population of US Tax Court Written Decisions*. Virginia Tax Review. Volume 30, Page 523 [↗](#)
- [23] **Bommarito II, M. J.**, & Katz, D. M. (2010). *A mathematical approach to the study of the United States Code*. Physica A: Statistical Mechanics and its Applications. Volume 389, Issue 19, Pages 4195-4200 [↗](#)

- [24] **Bommarito, M. J.**, Katz, D., & Zelner, J. (2009). *Law as a seamless web? Comparison of various network representations of the United States Supreme Court corpus (1791-2005)*. Proceedings of the 12th International Conference on Artificial Intelligence and Law. ICAIL 2009, pp. 234-235 ↗

Preprints, Reports & Working Papers

- [1] **Bommarito, M. J., II** (2026). *The Windows IOCTL Census: A Corpus-Scale, Multi-Architecture Database of the Driver Control-Code Surface*. Working paper (draft). *Draft manuscript*.
- [2] **Bommarito, M. J., II** (2026). *The Scrutiny Gradient: A Mean-Field Survey of Security Commits, CVEs, and Dossiers in 22 Linux Base-System Repositories*. Working paper (draft). *Draft manuscript*.
- [3] **Bommarito, M. J.** (2026). *OpenMPSC: An Open Dataset of Michigan Public Service Commission Regulatory Proceedings*. SSRN. Posted: April 2026 ↗
- [4] **Bommarito, M. J., II** (2026). *Needles at Scale: LLM-Assisted Target Selection for Windows Vulnerability Research*. arXiv preprint. arXiv preprint arXiv:2606.01364 ↗
- [5] **Bommarito, M. J.** (2026). *Moratorium Nation: A Survey of Data Center, Renewable Energy, and Battery Storage Moratoria in the United States*. SSRN. 113 Pages, Posted: 17 Feb 2026 ↗
- [6] **Bommarito, M. J.** (2026). *MimeLens: Pretrained Encoders for Fine-Grained Content-Type Detection*. Working paper (draft). *Draft manuscript*.
- [7] **Bommarito, M. J.** (2026). *ChunkLoris: Measuring Per-Chunk HTTP Amplification*. Working paper (draft). *Draft manuscript*.
- [8] **Bommarito II, M. J.**, Bommarito, J., & Katz, D. M. (2025). *The KL3M Data Project: Copyright-Clean Training Resources for Large Language Models*. arXiv preprint. arXiv preprint arXiv:2504.07854 ↗
- [9] **Bommarito, M. J.**, Katz, D. M., & Bommarito, J. (2025). *Precise Legal Sentence Boundary Detection for Retrieval at Scale: NUPunkt and CharBoundary*. arXiv preprint. arXiv preprint arXiv:2504.04131 ↗
- [10] **Bommarito, M. J.** (2025). *OpenGloss: A Synthetic Encyclopedic Dictionary and Semantic Knowledge Graph*. arXiv preprint. arXiv preprint arXiv:2511.18622 ↗
- [11] **Bommarito, M. J.**, Katz, D. M., & Bommarito, J. (2025). *KL3M Tokenizers: A Family of Domain-Specific and Character-Level Tokenizers for Legal, Financial, and Preprocessing Applications*. arXiv preprint. arXiv preprint arXiv:2503.17247 ↗
- [12] **Bommarito, M. J.** (2025). *Binary-30K: A Heterogeneous Dataset for Deep Learning in Binary Analysis and Malware Detection*. arXiv preprint. arXiv preprint arXiv:2511.22095 ↗
- [13] **Bommarito, M. J.** (2025). *Binary BPE: A Family of Cross-Platform Tokenizers for Binary Analysis*. arXiv preprint. arXiv preprint arXiv:2511.17573 ↗
- [14] Katz, D. M., Hartung, D., Gerlach, L., Jana, A., & **Bommarito II, M. J.** (2023). *Natural language processing in the legal domain*. arXiv preprint. arXiv:2302.12039 ↗
- [15] Bommarito, J., **Bommarito, M.**, Katz, D. M., & Katz, J. (2023). *GPT as knowledge worker: a zero-shot evaluation of (AI) CPA capabilities*. arXiv preprint. arXiv:2301.04408 ↗

- [16] **Bommarito II, M.**, & Katz, D. M. (2022). *GPT takes the bar exam*. arXiv preprint. arXiv:2212.14402 [↗](#)
- [17] **Bommarito, M. J.**, & Bommarito, E. (2021). *An empirical analysis of the R package ecosystem*. arXiv preprint. arXiv:2102.09904 [↗](#)
- [18] Bommarito, E., & **Bommarito, M.** (2019). *An empirical analysis of the Python Package Index (PyPI)*. arXiv preprint. arXiv:1907.11073 [↗](#)
- [19] **Bommarito, M. J.**, Katz, D. M., & Detterman, E. M. (2018). *OpenEDGAR: Open source software for SEC EDGAR analysis*. MIT Computational Law Report. arXiv:1806.04973 [↗](#)
- [20] Katz, D. M., **Bommarito II, M. J.**, & Blackman, J. (2017). *Crowdsourcing accurately and robustly predicts Supreme Court decisions*. arXiv preprint. arXiv:1712.03846 [↗](#)
- [21] Katz, D. M., & **Bommarito, M. J.** (2016). *Legal analytics course*. Course Materials.
- [22] Poast, P., **Bommarito, M. J.**, & Katz, D. M. (2015). *The Electronic World Treaty Index: Collecting the Population of International Agreements in the 20th Century*. SSRN. SSRN 2652760 [↗](#)
- [23] Katz, D. M., **Bommarito, M. J.**, Soellinger, T., & Chen, J. M. (2015). *Law on the market? Evaluating the securities market impact of Supreme Court decisions*. arXiv preprint. arXiv:1508.05751 [↗](#)
- [24] **Michael J. Bommarito II** (2011). *Law ? Computation*. University of Houston Law and Computation Workshop. Presentation at the University of Houston Law and Computation Workshop, Houston, TX
- [25] **Bommarito, M. J.** (2010). *The Race to the Bund: Aggregate European and State-Level Integration from the EEC to Present*. SSRN. [↗](#)
- [26] **Bommarito, M. J.** (2010). *Intraday Correlation Patterns between the S&P 500 and Sector Indices*. SSRN. [↗](#)
- [27] **Bommarito II, M. J.**, & Katz, D. M. (2010). *Building the United States Supreme Court Disposition Corpus 1791-2009*. Linguistic Data Consortium. [↗](#)
- [28] **Bommarito II, M. J.**, & Katz, D. M. (2009). *Properties of the United States Code citation network*. arXiv preprint. arXiv:0911.1751 [↗](#)
- [29] **Bommarito, M. J.** (2009). *Exploring Relationships between Legal Concepts in the United States Supreme Court*. SSRN. Available at SSRN 1814169 [↗](#)

Selected Open-Source Software, Datasets & Models

- **OpenMPSC Data** [Dataset] — Open dataset of Michigan Public Service Commission regulatory proceedings, 1987–2026, with a public REST API at openmpsc.com [↗](#)
- **linux-drivers.com** [Dataset] — Evidence-backed dossiers for 864 Linux kernel drivers, recommending which legacy drivers to keep, annotate, deprecate, or remove [↗](#)
- **Michigan Energy Infrastructure & Population Density Map** [Dataset] — Multi-layer geospatial visualization of Michigan’s energy infrastructure overlaid on Census population density, with a data center siting score for all 1,580 county subdivisions

- **OpenGloss** [Software] — A synthetic encyclopedic dictionary and semantic knowledge graph for English with 537K sense definitions, 9.1M semantic edges, and 60M words of encyclopedic content ↗
- **NUPunkt** [Software] — High-precision sentence boundary detection for legal text ↗
- **KL3M Tokenizers** [Software] — Domain-specific BPE tokenizers for legal, financial, and governmental text, achieving 9-17% efficiency improvements over GPT-4 and LLaMA3 ↗
- **KL3M Data Project** [Dataset] — Large-scale copyright-clean dataset containing 132M+ documents and trillions of tokens for training legal language models ↗
- **FOLIO - Federated Open Legal Information Ontology** [Dataset] — Open-source legal data standard containing 18,000+ standardized legal concepts with multilingual support for improved legal industry interoperability ↗
- **LexNLP** [Software] — Natural language processing and information extraction for legal and regulatory text ↗
- **OpenEDGAR** [Software] — Open source Python client for SEC EDGAR data access and analysis ↗

Honors & Distinctions

Fastcase 50	2019
NSF Research Assistantship, Center for the Study of Complex Systems	2010–2011
NSF Fellowship, IDEAS-IGERT	2009–2010
NSF Research Assistantship, School of Natural Resources & Environment	2008–2009
Google Summer of Code (participant)	2007, 2008, 2009