EXPERIMENTAL JURISPRUDENCE

Will Henson*

TABLE OF CONTENTS

INTRODUCTION

The intersection of artificial intelligence (AI) and legal analysis can revolutionize how we address foundational questions in legal philosophy. This Technology Explainer discusses how artificial intelligence might address the biggest problems in legal analysis on an unprecedented scale. Experimental jurisprudence is an emerging field that combines legal theory and empirical methods to make progress on fundamental questions of legal philosophy.[1] Drawing from cognitive and behavioral science[2] among other disciplines,[3] it focuses on the *interpretation* of data to make progress on legal philosophy.[4]

---

* Georgetown University Law Center, J.D. Candidate 2024; University of Alabama, B.A. Theater & Philosophy, 2021.

[1] *See* Kevin Tobia, *Experimental Jurisprudence*, 89 U. CHI. L. REV. 735 (2022).

[2] *See* Roseanna Sommers, *Experimental Jurisprudence*, 373 SCI. 394 (2021).

[3] Tobia, *supra* note 1, at 779 (explaining how experimental jurisprudence might apply corpus linguistics, the processing of large bodies of text. Surveying laypeople is the most common method in experimental jurisprudence, but the responses collected can also be used as a "corpus" for *linguistic* analysis in addition to the common statistical analyses).

[4] *See id.* (explaining that while experimental jurisprudence often relies on the creation of data, a rich body of research in the field has turned to the processing and interpretation of existing data).

Machine learning (ML), an application of artificial intelligence (AI) capabilities to process large amounts of data, has arisen as a potential tool to help. Applied to experimental jurisprudence, ML treats judicial opinions as data[5] to identify trends, such as in the use of interpretive tools. For example, ML can be used to understand the rise of textualism, an interpretive tool that prioritizes the semantic language of a statute, ostensibly by considering how a reasonable person would understand it, and how it is applied.[6] By the same token, ML can also be used to analyze legal texts for things other than interpretive tools. For example, intellectual property scholars might be interested in how various jurisdictions analyze the likelihood of confusion test in trademark cases. ML can then be used to discover patterns over the weight judges accord to each element.[7]

In this Technology Explainer, I illustrate how legal scholars in this emerging field can use AI to analyze massive quantities of judicial opinions to identify trends within and between their pages. To do so, I define and explain the most basic steps of this methodology: labeling, pre-processing, and pre-training. Finally, I conclude with an appeal to advocating scholars interested in the field to make use of this methodology.

I. USING AI TO ANALYZE JUDICIAL OPINIONS

AI generally encompasses a wide variety of techniques used to approximate human intelligence using machines.[8] Much of the progress in AI today comes from a subset of these techniques: machine learning.[9] ML is the programming of a system to improve

---

[5] Big data is the collection of increasingly massive quantities of information, often enumerated in file size rather than number. Francis X. Diebold, *On the Origin(s) and Development of "Big Data": The Phenomenon, the Term, and the Discipline* 7 (PIER Working Paper No. 12-037, 2012) ("[I]n a landscape littered with failed attempts at interdisciplinary collaboration, Big Data is emerging as a major interdisciplinary triumph").

[6] Anita S. Krishnakumar, *Statutory Interpretation in the Robert's Court's First Era: An Empirical and Doctrinal Analysis*, 62 HASTINGS L.J. 221, 223 n.2 (2010) ("Textualism is an interpretive philosophy that prioritizes the statute's text above all else."); John F. Manning, *What Divides Textualists from Purposivists?*, 106 COLUM. L. REV. 70, 76 (2006).

[7] *See, e.g.*, Polaroid Corp. v. Polarad Elecs. Corp., 287 F.2d 492 (2d Cir. 1961).

[8] Ryan Calo, *Artificial Intelligence Policy: A Primer and Roadmap*, 51 U.C. DAVIS L. REV. 399, 404 (2017).

[9] *Id.*

its performance over time, often by identifying patterns in datasets.[10] This is where big data comes into play. The ML application processes massive amounts of information and learns from the patterns therein, identifying patterns that many people cannot.[11] Commentators, including Supreme Court justices themselves, have noted that textualism is on the rise.[12] This is ultimately an empirical claim, and is therefore perfect for the attention of ML and experimental jurisprudence.[13] This Part illustrates how AI can be used to process judicial opinions and generate legally significant finding, and uses as an example a study by Austin Peters that leveraged ML to analyze thousands of state Supreme Court opinions dating back to 1980 to track the prevalence of textualism over time.[14]

A. Labeling Data for Supervised ML Model

The Peters study, and others like it,[15] opted for a supervised ML model.[16] In supervised models, each piece of relevant training data must be labeled before the ML model is trained.[17] This labeled data is essential for the model to learn patterns and make predictions.[18]

In particular, the data is labeled so that it can serve as training data to teach an ML system how to classify future cases in the same way. It builds a map from inputs to outputs so that the algorithm can then generalize, or be applied more broadly to, unseen, unlabeled

---

[10] *Id.* at 405.

[11] *Id.* at 421 ("[T]he *entire purpose* of AI is to spot patterns people cannot").

[12] Harvard Law School, *The Antonin Scalia Lecture Series: A Dialogue with Justice Elena Kagan on the Reading of Statutes*, YOUTUBE (Nov. 25, 2015), https://www.youtube.com/watch?v=dpEtszFT0Tg [https://perma.cc/428L-G4B7].

[13] Tobia, *supra* note 1, at 783.

[14] Austin Peters, Are They All Textualists Now? (2023) (Ph.D. dissertation, Stanford University) (on file with author).

[15] *See* BRANDON M. STEWART, MARGARET E. ROBERTS & JUSTIN GRIMMER, TEXT AS DATA: A NEW FRAMEWORK FOR MACHINE LEARNING AND THE SOCIAL SCIENCES 178–83 (2022).

[16] *See* Peters, *supra* note 14, at 8.

[17] Timothy Lee & Sean Trott, Large Language Models, Explained with a Minimum of Math and Jargon, Understanding AI (July 27, 2023), https://www.understandingai.org/p/large-language-models-explained-with [https://perma.cc/RJN4-NJK2].

[18] Iqbal H. Sarker, *Machine Learning: Algorithms, Real-World Applications and Research Directions*, 2 SN COMPUT. SCI. 160, 162 (2021) ("[Supervised machine learning] uses labeled training data and a collection of training examples to infer a function.").

data.[19] The learning algorithm is provided with a set of input-output pairs, and it learns from this data to carry out a task, such as make predictions, classifications, or decisions, when new data is encountered. In a sense, the labeled data teaches the algorithm how to label its *own* cases at scale.[20]

To illustrate a supervised ML, imagine you were handed a collection of someone's family photos and tasked with completing a scrapbook with names written underneath. To match the names properly, it would be critically important for you to have some pictures with each member of the family properly identified. Without this set of labeled photos, assigning names to each member of the family would return random results.[21] However, with the labeled data as a starting point, you would be able to match each new picture to one from the labeled set and get a reasonably accurate categorization.

In unsupervised learning, the algorithm is given only input data without corresponding output labels, with the objective of finding hidden patterns or structures within the data.[22] Unlike supervised learning, there are no teacher-labels to provide guidance, so the algorithm must find relationships or groupings on its own.[23] Unsupervised learning is often used for tasks such as clustering, where the algorithm attempts to group similar examples together.[24]

Returning to the scrapbooking example, unsupervised ML would be analogous to if your friend did not give you any labeled photos to begin with, and instead asked you to simply group the people in the photos by physical similarity. You may be able to successfully group photos of the same people together, and you may even be able to distinguish which branch of the family different groups belong to based on their shared features. However, you would lack any frame of reference for *who* these people were, and the larger the family grouping (cluster) became, the more likely it would be that you placed a person in the wrong group, for example, a maternal cousin in the paternal cousin group.

Supervised ML models are better suited to the analysis of judicial opinions where scholars are interested in particular trends.

---

[19] CHRISTOPHER M. BISHOP, PATTERN RECOGNITION AND MACHINE LEARNING 3 (2006).

[20] Sarker, *supra* note 18, at 162.

[21] You may make some guesses based on gender or age, but the results would be far from perfect.

[22] BISHOP, *supra* note 19, at 3.

[23] *Id.*

[24] *Id.* Other uses of unsupervised learnings that are less relevant for the purposes of this paper are density estimation, which is determining the distribution of the data among inputs, and visualization, which is simplifying data into two or three dimensions for use in graphs.

While unsupervised ML may be useful for scholars interested in discovering "associations" between the reasoning of various opinions or "clustering" similarly-written holdings together, it cannot classify the opinions by methodological tools nor incorporate context.[25] The fundamental issue of unsupervised learning is that it lacks inherent direction and goals can be difficult to define because it merely *identifies* patterns, rather than looking for specific patterns.[26] Furthermore, supervised learning can benefit from the labels already built in to legal databases such as Westlaw's headnotes.[27]

In Peters's study, he began the labeling process for supervised learning by taking 44,000 state supreme court cases listed in Westlaw under the headnote for "statutory interpretation" and took a random sample of 3,000 cases from that initial set.[28] To create the actual labels, Peters further cut these cases down to specific *paragraphs* about statutory interpretation[29] and labeled the paragraph with a "1" for textualist tools (although any random marker could be used).[30] The specific tools that Peters marked were plain meaning, references to the dictionary, or the use of a linguistic canon.[31] These tools of textualism represent what the ML system will focus on.

Textualism can also be understood by its *rejection* of other tools that are seen as external to the text of the statute, like legislative

---

[25] Julianna Delua, *Supervised vs. Unsupervised Learning: What's the Difference?*, IBM (Mar. 12, 2021), https://www.ibm.com/blog/supervised-vs-unsupervised-learning/ [https://perma.cc/C93K-XVWD].

[26] Etienne Bernard, *Clustering*, IN INTRODUCTION TO MACHINE LEARNING(2021) (ebook).

[27] *Id.* (the main difference between supervised and unsupervised learning is the availability and use of labeled data).

[28] Peters, *supra* note 14, at 13 (This keyword, or "dictionary," approach is also an example of labeled data; instead of Peters labeling them by hand, Westlaw had labeled these 44,000 cases for their statutory interpretation content. Furthermore, instead of running a machine learning algorithm to learn from *this* label, Peters simply applied an internal heuristic by eliminating any cases from the set that did not have this statutory interpretation label.).

[29] *Id.* at 12 ("Technically, it is easier for machine learning models to make predictions about paragraphs than entire opinions since these text chunks are smaller.").

[30] *See id.* at 13 (identifying three textualist tools of statutory interpretation: (1) use of plain meaning rule, (2) reference to the dictionary, or (3) use of a linguistic canon).

[31] Peters, *supra* note 14, at 13.

history.[32] If the paragraph contained no textualist tool or simply referenced a textualist tool (but did not apply it), then Peters labeled the paragraph "0." This represents an advantage over simple keyword searching, which tends to count paragraphs that *describe* a textualist tool but ultimately rejects it, leading to false positives.[33] This is because keyword searches only search for single words or word pairs in a text and treats all hits in all parts of an opinion equally, ignoring the broader context.[34] Returning back to the scrapbooking labeling example, this sort of supervised labeling process would be akin to hand-labelling 300 photographs with the names of family members, in the hopes this would help the ML system label the remaining 4,400 photographs.[35]

B. Pre-Processing the Data

The next step is pre-processing, or cleaning the raw data before it can be used for ML.[36] Cleaning the data, otherwise known as normalization, is the process of reducing redundancy and improving consistency in a database by applying normal forms to inputs and tables.[37] Pre-processing can take many forms, but it often includes the removal of word tense endings (like -ing or -ed), deleting the -s from the end of plurals, and deleting articles and conjunctions.[38] This process can improve performance and reduce the number of "tokens" (parts of words and punctuation symbols) that the algorithm has to

---

[32] CROSS, *supra* note 15 ("Modern textualism essentially rests on a 'common core of rejecting legislative history.'").

[33] *See* Peters, *supra* note 14, at 10 (such an approach would similarly generate false negatives by failing to label paragraphs that used textualist tools but did not call them by name).

[34] *Id*; *see* Pablo Barberá, Amber E. Boydstun, Suzanna Linn, Ryan McMahon & Jonathan Nagler, *Automated Text Classification of News Articles: A Practical Guide*, 29 POL. ANALYSIS 19, 32 (2021) ("[N]ot all features in the corpus and relevant to the analysis will be in the dictionary."); *see also* STEWART, ROBERTS & GRIMMER, *supra* note 15, at 181 (emphasizing the importance of context in corpus analysis).

[35] In this study, however, it is even simpler than the analogy suggests. Because Peters is labeling paragraphs as "1" = textualism and "0" = no textualism, it would be more accurate to compare the process to labeling the remaining photographs for "Dad" or "no Dad" (rather than every member of the family). The additional names might help you identify what "no Dad" means, but the task could probably be completed with only pictures of "Dad."

[36] STEWART, ROBERTS & GRIMMER, *supra* note 15, at 52–55.

[37] E.F. Codd, *A Relational Model of Data for Large Shared Data Banks*, 13 COMMS. OF ACM 377, 377 (1970).

[38] *Id.*

deal with.[39] Generally, the more tokens that an ML system has to process, the more computing power it will require, and the more likely that the "noise" of unnecessary tokens will distort the results ("garbage in, garbage out").[40] Similar to human memory, the ability of an ML system to process relevant information depends on the number of "tokens" it has to contend with.[41] Especially when dealing with smaller datasets (such as judicial opinions, which come in the thousands rather than millions or billions), each additional piece of unclean data can have an outsized impact, as it represents a greater portion of the overall set.[42]

This process is similar to preparing ingredients before cooking a dish. Just as spoiled parts must be removed from vegetables, to clean his data set for the ML, Peters removed "stop words" (common words with little semantic meaning) and numbers from the case text.[43] Removing the unnecessary elements of ingredients before cooking can speed up the process, since the cook no longer has to interrupt their workflow to deal with extraneous components. Peters also normalized the entire dataset by transforming the paragraphs to word tokens and stemming words (removing suffixes or prefixes to transform them into their root form).[44] This is analogous to cutting ingredients into similar sizes for even cooking, because it puts all the data features into a similar format that is easier to read or process.

---

[39] STEWART, ROBERTS & GRIMMER, *supra* note 15, at 52–55 (listing strategies for cleaning data before training).

[40] Samadrita Ghosh, *A Comprehensive Guide to Data Preprocessing*, NEPTUNE.AI (Aug. 22, 2023), https://neptune.ai/blog/data-preprocessing-guide [https://perma.cc/XQ6A-LL3R].

[41] Jonathan H. Choi, *How to Use Large Language Models for Empirical Legal Research*, (Aug. 13, 2023) J. INST. & THEORETICAL ECON. (forthcoming 2024) (manuscript at 5).

[42] *See* David Lehr & Paul Ohm, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, 51 U.C. DAVIS L. REV. 653, 681 (2017).

[43] Peters, *supra* note 14, at 14 .

[44] *Id.*

C. Pre-Training the Data

The next step in the process is the pre-training of the ML model.[45] The goal here is to develop the capabilities of the model.[46] Models in legal scholarship are likely to be built on text, and this step helps develop the model's ability to classify portions of text or discover patterns in semantic meaning.[47] This stage generates the design of the model itself that can be deployed to accept different input data and generate different output data.[48] Pre-training is most useful for relatively smaller datasets, such as judicial opinions.[49]

The first step is to define a pre-training objective that leverages the labeled and pre-processed data. For classification of legal corpus, the objective might be to predict the interpretive method, topic, or sentiment of different paragraphs. The ML system then takes that objective (here, a classification of interpretive method) to try to produce accurate outputs with the inputs it is given.[50] At first, it will make random guesses about which inputs (e.g., paragraphs) match with the right outputs (e.g., textualist or non-textualist), and may get many of those guesses wrong.[51]

Where machine *learning* comes in is when that system takes the labeled paragraphs and uses them to check itself, essentially "learning" which guesses are right and which are wrong. This is known as the "training" set.[52] It calculates the difference between actual and predicted inputs, known as the "error," and repeats this guess and check process until it can reliably generate accurate outputs.[53] For the human researcher to determine when the desired level of accuracy has been achieved, they may take a much smaller

---

[45] Paul Ohm, *Focusing on Fine-Tuning: Understanding the Four Pathways for Shaping Generative AI*, COLUM. SCI. & TECH. L. REV. (forthcoming 2024) https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4738261 [https://perma.cc/547A-9P8J] (defining the initial training process as "pretraining" because two other steps, "fine-tuning" and "in-context learning," can also be used to train the data after the initial model is produced).

[46] *Id.* at 8.

[47] *Id.*

[48] *Id.* at 10.

[49] Jeremy Howard & Sebastian Ruder, *Universal Language Model Fine-tuning for Text Classification*, 1 56TH ANN. MEETING ASSOC. COMP. LING. 238, 344 (2018).

[50] Doug Rose, *What Are Artificial Neural Networks?*, *in* ARTIFICIAL INTELLIGENCE FOR BUSINESS (2020) (ebook).

[51] *Id.*

[52] *Id.*

[53] *Id.*

set of labeled data that was segregated from the training set, called the "test" set, and see how the model predicts outputs for data that it did not directly learn from.[54]

Returning to the scrapbooking example, this process is analogous to a person hand-labeling 300 of their family photos but only giving you 240 to learn from. Then, to make sure you had a good handle on the project before turning you loose on the remaining 4,400 photographs, the person asks you to self-check your ability to match faces with names on the remaining 60 hand-labeled photographs.

In the Peters study, the author randomly selected 80% of the hand-labeled paragraphs to be the training set and 20% to be the test set.[55] By training an algorithm on the former set, the latter set can be used to evaluate the accuracy of the ML model.[56] "Since the model never 'sees' the test set, [the performance with the] test set sheds light on the model's performance for out-of-sample paragraphs."[57] In his study, Peters programmed the algorithm to predict whether a given paragraph used textualism.[58] The model predicted textualism with an accuracy of 94%.[59]

II. WHY LEGAL SCHOLARS SHOULD EMPLOY ML

Legal scholars, practitioners, and judges are already making use of ML. One study used similar ML systems on the text of small-claims trial court decisions, which were able to predict the outcome of appeals in Brazilian courts better than human experts.[60] A forthcoming study purports to use the OpenAI large language model, a specific application of ML, to classify judicial passages for their use of other canons of statutory interpretation, like the rule against surplusage ("one of the most important and frequently cited canons of construction") more efficiently and at greater scale than human research assistants.[61] Outside of academic literature, ML has taken

---

[54] *Id*.

[55] Peters, *supra* note 14, at 10.

[56] *Id*. at 15.

[57] *Id.*

[58] *Id*. at 16 (first analyzing hand-labeled paragraphs in the test set before applying the methodology to the remaining 41,000 out-of-sample paragraphs).

[59] *Id*.

[60] Elias Jacob de Menezes-Neto & Marco Bruno Miranda Clementino, *Using deep learning to predict outcomes of legal appeals better than human experts: A study with data from Brazilian federal courts*, 17 PLOS ONE 7, 17–18 (2022).

[61] Choi, *supra* note 41, at 3.

an even greater role in the law by guiding government officials in their decisions.[62] Jurisdictions across the U.S., from local municipalities to the federal government, have adopted ML platforms to assist in all manner of tasks such as predicting where crime will occur, distributing welfare, detecting tax fraud, and setting bail.[63] Given the proliferation of use of ML in the legal system, it reasons that legal theorists adopt it to *analyze* those very same systems.

First, such broad empirical findings are only possible with data-processing tools like ML.[64] Before ML, empirical studies of judicial opinions were restricted in time frame and/or scope because they relied *solely* on hand-coding methods.[65] Rather than speculate on trends in judicial decision-making, experimental jurisprudence allows scholars to identify those trends. Such an approach may illuminate data on trends across time, jurisdiction, or subject matter.[66] For example, Peters found that the use of textualism increased by 84% from 1980 to 2019 across tens of thousands of state Supreme Court opinions and across the full ideological spectrum.[67]

Some critics of this approach argue that the use of ML is not a legitimate legal philosophy.[68] They argue that studying the language of opinions in this format cannot truly express what judges were *thinking*, and that the written opinion is simply "window dressing" that obscures the judges' true, internal beliefs about a given case.[69] However, the written opinion, not the judges' internal beliefs, is what becomes law and binds future judges by *stare decisis*.[70] The written

---

[62] This is not necessarily a good thing, but it is hard to understate the adoption of ML in government. *See* Ben Green, *The Flaws of Policies Requiring Human Oversight of Government Algorithms*, 45 COMP. L. & SEC. REV. 105681 (2022).

[63] Aziz Z. Huq, *Artificial Intelligence and the Rule of Law*, 764 PUBLIC LAW AND LEGAL THEORY, 1, 3–4 (2021).

[64] Peters, *supra* note 14, at 11.

[65] *See, e.g.*, Anita S. Krishnakumar, *Statutory Interpretation in the Roberts Court's First Era: An Empirical and Doctrinal Analysis,* 62 HASTINGS L.J. 221, 231 (2010) (hand coding Supreme Court cases from 2005–2008).

[66] Tobia, *supra* note 1, at 783.

[67] Peters, *supra* note 14, at 59.

[68] Herman Cappelen, *Précis of Philosophy Without Intuitions*, 171 PHIL. STUD. 513, 513–15 (2014) ("So the worry isn't just that a few meta philosophers have some false beliefs about how philosophy is practiced.").

[69] Amy Semet, *An Empirical Examination of Agency Statutory Interpretation*, 103 MINN. L. REV. 2255, 2286 (2019).

[70] Antonin Scalia, *Common-Law Courts in a Civil-Law System: The Role of United States Federal Courts in Interpreting the Constitutional*

opinion is the lodestar for advocates as they develop their cases in any given jurisdiction. Even advocates who attempt to get inside the judges head, so to speak, can look no further than the written opinion. Likewise, since ML can only study observable phenomena, the written opinion seems an excellent place to start: you cannot manage what you cannot measure. This limitation applies to any scholarship that analyzes legal texts, and taking this critique to its logical conclusion would mean that such study is never meaningful.[71] Rather than abandon an entire body of jurisprudence, it is probably more realistic to acknowledge this limitation and continue to focus on the only external expression of the judges' beliefs that is available: the text itself.

Finally, while no ML model is 100% accurate, and the training set composed of opinions is minuscule compared to systems like ChatGPT,[72] the hybrid approach of supervised ML is better able to eliminate both false positives and false negatives than a simple keyword search. To the extent that errors *do* exist, researchers can quantify them and be transparent about their frequency.[73] This method also has benefits over unsupervised ML, which follows a similar procedure, but without labeled data.[74] While it may save time by avoiding the lengthy labeling process, unsupervised ML is better at finding patterns *between* datapoints rather than between the data and the label.[75] It may be possible for scholars to apply unsupervised learning to discover hidden patterns in the judicial corpus and apply labels ex post, but there are two problems with this approach: (1) the unsupervised clustering could not answer questions about specific trends in the data as it eschews the use of labels by definition and without labels, an ML system lacks direction,[76] and (2) this hybrid human-AI approach would risk serious inaccuracy in the interpretation of data as it would require a human to make assumptions about the "logic" of the algorithm (meaning that the label applied ex post might not actually be the grouping found by the

---

Law, TANNER LECTURES ON HUMAN VALUES 92 (1995) ("It is the *law* that governs, not the intent of the lawgiver.").

[71] Peters, *supra* note 15, at 32 n.161.

[72] Ohm, *supra* note 45, at 9 (explaining that GPT-4 may have been trained on a petabyte of data).

[73] Peters, *supra* note 14.

[74] Sarker, *supra* note 18, at 163 ("Unsupervised learning analyzes unlabeled datasets without the need for human interference").

[75] *Id*. at 167 ("Cluster analysis… is an unsupervised machine learning technique for identifying and grouping related data points in large datasets without concern for the specific outcome.").

[76] Bernard, *supra* note 26.

ML system).[77] In other words, it does not illuminate how information fits in a broader context (like judicial interpretative trends) as effectively as supervised ML. Were this study performed with unsupervised ML, the code may be able to cluster different types of judicial philosophy together to discover interesting trends,[78] but it would not be able to say whether a given opinion was textualist or non-textualist, and the trends it could predict would be unmoored from the ideology that is of interest to legal scholars. Furthermore, unsupervised learning may require even more data than supervised learning to generate accurate results.[79] Unsupervised models can overfit the data and draw associations that do not truly exist; to address this risk, more and more data is necessary to generate accurate results.[80] Because there is a limited number of judicial opinions to draw from, supervised ML is a more efficient use of resources (even if it takes more time to label the data).[81] There were only 44,000 state Supreme Court cases on point in the Peters example; likewise, there are only 30,863 U.S. Supreme Court opinions ever issued.[82] Compare this to the best-known unsupervised ML model, Chat-GPT, which was trained on 300 billion individual inputs.[83]

---

[77] Green, *supra* note 62, at 105687 ("A long-standing body of research shows that, across a wide range of domains, automated decision-support systems tend to alter human decision-making in unexpected and harmful ways.").

[78] Sarker, *supra* note 18, at 168 ("[Unsupervised learning] is often used as a data analysis technique to discover interesting trends or patterns in data").

[79] Muhammad Usama, Junald Qadir, Aunn Raza & Hunain Arif., *Unsupervised Machine Learning for Networking: Techniques, Applications and Research Challenges*, 7 IEEE ACCESS 65579, 65606 (2019) ("[D]ue to the unavailability of labels or related information, unsupervised learning models can overfit the data, which causes issues in testing").

[80] *Id.*

[81] *Cf.* Peter David Turney, *Types of Cost in Inductive Concept Learning*, PROC. COST-SENSITIVE WORKSHOP AT THE 17TH ICML-2000 CONF. (2000) (explaining the unmeasurable human cost of labeling data including the time it takes and the "bottleneck" of finding a domain expert to correctly label it; here, there is a similar "bottleneck" because labelers require a minimum of legal knowledge to correctly identify textualism in judicial opinions).

[82] *Supreme Court Database*, WASH. UNIV. L. (Dec. 4, 2023), http://supremecourtdatabase.org/ [https://perma.cc/X565-2EZZ].

[83] Beatrice Nolan, *Google researchers say they got OpenAI's ChatGPT to reveal some of its training data with just one word*, BUS. INSIDER (Dec. 4, 2023), https://www.businessinsider.com/google-researchers-openai-

CONCLUSION

While experimental jurisprudence is imminently capable of finding useful results like Peters did here, the methods can be arcane to a layperson or even to other legal scholars with limited backgrounds in machine learning. However, when the process is broken down into individual components, it is easier to see how practitioners such as Peters arrived at their conclusions, and hopefully the methods can be adopted for future work by those interested in applying cutting-edge techniques to reach empirical results. When it comes to application of this methodology, the sky is the limit; legislative history, public financial disclosures, agency proceedings, and, of course, judicial decisions are all data-rich corpuses with trends that could be illuminated by machine learning.

chatgpt-to-reveal-its-training-data-study-2023-12#:~:text=The%20AI%20model%20powering%20ChatGPT,or%20570%20GB%2C%20of%20data. [https://perma.cc/FZ22-PXF9].