Title

Essays on Law and Natural Language Processing

Author

Akdemir, Elez Ilya

Publication Date

2023

Peer reviewed|Thesis/dissertation

Essays on Law and Natural Language Processing

By

Elez Ilya Akdemir

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of the Science of Law – Juris Scientiae Doctor (J.S.D.)

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Kenneth Ayotte, Chair
Professor Adam Badawi
Professor David Bamman

Summer 2023

Essays on Law and Natural Language Processing

Abstract

Essays on Law and Natural Language Processing

By

Elez Ilya Akdemir

Doctor of the Science of Law – Juris Scientiae Doctor (J.S.D.)

University of California, Berkeley

Professor Kenneth Ayotte, Chair

The present work, titled "*Essays on Law and Natural Language Processing*", explores the uses of Natural Language Processing (NLP) in law. Structurally, the first two essays explore the *applications* of NLP methods in the legal domain. The third essay explores the *epistemological issues* concerning NLP applications in the legal domain, particularly in the context of the possibilities of using large language model (LLM)-based artificial intelligence (AI) for legal decision-making. Accordingly, the first essay proposes the application of text classification methodology to the question of US Supreme Court's certiorari petition outcome prediction. The second essay proposes the application of word embedding-based synchronic lexical semantic change detection methodology to investigate a question in corporate governance regarding corporate honesty on Environmental, Social and Governance (ESG) issues. The third essay argues that while large language models are not explicitly designed to capture reasoning, they will likely exhibit latent legal epistemological biases rooted in Jean-Michel Berthelot's schemas of intelligibility-based patterns of legal reasoning detected in the text. Overall, this dissertation argues that the complex nature of legal reasoning and the resulting textual data generation necessitate not only a cautious approach to the application of NLP methods in law but also a re-examination of the epistemological foundations underpinning legal reasoning.

Scire leges non hoc est verba earum tenere,
sed vim ac potestatem.

[Knowing laws is not a matter of sticking to their words,
but a matter of grasping their force and tendency.]

Celsus, (1st Century AD)[1]

[1] Digest of Roman Law 1.3.17 (533 AD),
translated in Watson, Alan, ed. *The Digest of Justinian*. University of Pennsylvania Press, 2011.
Note: The Digest of Emperor Justinian is a compilation of earlier work by Roman Jurists. The quote is thus much older than 533 AD, as the jurist Celsus himself lived during the reign of Emperor Hadrian in the 1st Century AD.

To my family

Table of contents

Acknowledgments

First and foremost, I extend immense gratitude to my advisor and mentor, Professor Kenneth Ayotte, whose constant support, unwavering guidance, intellectual rigor, steadfast leadership and patient mentorship have been the absolute bedrock of this project. Your generosity, your dedication to nurturing both academic and personal growth, your incredible wealth of knowledge and your willingness to engage in in-depth intellectual discussions, in addition to your insightful constructive feedback have played an indispensable role in shaping the trajectory and quality of this research and have significantly enhanced my journey at Berkeley, making it an invaluable and enriching experience.

I am also enormously indebted to the members of my dissertation committee, Professor Adam Badawi and Professor David Bamman. Your scholarly perspectives, your valuable insights, your wealth of expertise and your discerning assessments have played a key role in honing the quality of my research. Moreover, your steadfast support and encouraging assistance have been indispensable in bringing this dissertation to its culmination.

I would also like to express gratitude to Professor Jonathan Marshall who has provided incredible support, guidance and mentorship throughout my studies at Berkeley. Being your GSI for LS100, LS123 and LS190, and many other courses, has, without a doubt, provided me with the most stimulating and enriching academic environment I could ever hope for. Having been guided by your wisdom over these years, as well as by the exceptional intellect of Berkeley students I've had the privilege of teaching in these courses, has truly been an extraordinary blessing.

My deep gratitude extends to Professor Kevin Quinn, whose influence has been pivotal in my academic journey. Upon completing your incredible statistics course, you recommended that I venture into the realm of natural language processing. The significance of your advice cannot be overstated for it has profoundly influenced both the direction of my research as well as the trajectory of my life, and thus I am incredibly thankful for the introduction to NLP that you facilitated. Despite being in Michigan, your support has endured throughout these years and your encouragement and guidance remain a steadfast foundation for my work, playing a central role in aiding me along this path.

I also wish to express my heartfelt gratitude to the esteemed professors at the University of Kent, my alma mater, where my venture into the world of law commenced. Notably, I extend my gratitude to Professor William Howarth for his invaluable guidance and support. Your insightful and memorable instruction in Environmental Law has left an enduring mark on my legal academic journey, setting an impressively high standard for

both legal scholarship and pedagogical excellence. I am also equally thankful to Professor Geoffrey Samuel for his captivating European Legal Systems class, which ranks among the most enlightening courses I have ever had the privilege of taking. Your thought-provoking and unforgettable instruction has instilled in me a profound interest in the subject matter of this dissertation, and the thought that Leibniz aspired to mathematize law has remained with me since its introduction in your course, akin to a vivid and enduring ever-recurring dream.

I extend special gratitude to the remarkable individuals who comprised the D-Lab's Computational Text Analysis Working Group, specifically Adam Anderson, Professor Caroline Le Pennec-Çaldichoury, Professor Ben Gebre-Medhin, and Jaren Haber. My involvement with this group has unequivocally served as an invaluable wellspring of distinctive NLP insights, substantially elevating and enriching my academic journey. The multitude of discussions, conversations, presentations, comments, and debates within this group have nurtured a profound and unparalleled sense of camaraderie and friendship.

Indeed, one of the highlights of any journey are the friendships we make along the way. During my time at Berkeley, fate granted me the privilege of sharing this bond with the amazing squad made up of Keegan Freiburger, Andre Sanchez, and Trevor Chistolini. Your presence in my life has been a source of immense joy and delight and my thanks go out to each of you for everything, and I hope our bond will endure throughout our lives. For indeed, hope is a bird with feathers.

Last but certainly not least, I am infinitely thankful, indebted and grateful to my family: to my mother Gabriella Aigul, my brother Afiny, my nephew Jordan and my niece Danielle, and my two sisters, Ivy and Erika. Thank you for your immeasurable and unconditional love, encouragement, empathy and understanding, for your belief in me, and for your constant selfless support and sacrifices which have paved the way for my accomplishments. Without you I am nothing, and with you I am everything. What I owe you is beyond evaluation.

1 Predicting the Outcome of Certiorari Petitions: a Text Classification Approach

Abstract

Traditionally, the empirical study of United States Supreme Court's decisions to grant or deny certiorari petition review has focused on the presence of cues as predictors of review. Advances in natural language processing and machine learning allow us to approach the problem from a text classification paradigm. The US Supreme Court's case selection process can thus be modelled as a binary text classification task where each certiorari petition is labeled as either granted or denied. Utilizing fourteenth amendment certiorari petitions' extracted textual content we investigate whether certain legal issues under the amendment can be predictive of review. To the best of our knowledge this is the first work applying text classification methods to the study of certiorari petition outcome prediction.

Quod crebro videt, non miratur,
etiam si cur fiat nescit.

[A common happening does not astonish,
even though the cause is unknown.]

Cicero, *De Divinatione* (44 BC)[2]

[2] Cicero, Marcus Tullius, *On Divination*, Book II, 22 (44 BC)
English translation from Pascal, Blaise. "*Pensées*" ed and translated by Roger Ariew (Hackett, 2004), p 209.

1.1 Introduction

A Petition for Writ of Certiorari, or cert. petition for short, is a legal document submitted by a losing party requesting the review of a lower court's final judgement by the United States Supreme Court. A granted certiorari petition leads to a full hearing of the case by the Supreme Court, after which the Justices of the Supreme Court write their opinions on the legal question at hand. Conversely, a denied certiorari petition means that the final judgement of the court below stands, which, for all intents and purposes, puts an end to the case in question.[3] A fundamental characteristic of certiorari petitions is that only a small number of these petitions are granted review by the US Supreme Court. For instance, in 2013, out of 7326 certiorari petitions filed with the Supreme Court, only 76 were accepted, or roughly a 1% acceptance rate.

Granted certiorari petitions play an important agenda-setting role for the Supreme Court and ultimately result in written decisions on critical legal issues of the day.[4] But denied petitions also contain vital information, and as Linzer points out: "many times [a certiorari petition] denial gives us a glimpse, imperfect to be sure, into the Justices' preliminary attitudes on a given issue."[5]

Given the immense impact of US Supreme Court's decisions on America's political, social, and legal landscapes, a great deal of research has been dedicated to the study of why some petitions are granted review while others are denied. In this work I propose a novel methodology based on natural language processing and supervised machine learning approaches, namely, the adoption of a binary text classification approach to the question US Supreme Court's certiorari petition review. It is argued that a binary text classification-based approach can allow one to explore and take into consideration the so-far neglected aspect of certiorari petition outcome prediction research, namely the textual content of certiorari petitions themselves.

[3] Perry, Hersel W. Deciding to decide: agenda setting in the United States Supreme Court. Harvard University Press, 2009.
[4] *Ibid.*
[5] Linzer, Peter. "The meaning of certiorari denials." Colum L. Rev. 79 (1979): 1227.

In the first part of the present work, I introduce the relevant US Supreme Court procedure concerning certiorari review as a basis for exploring certiorari outcome prediction. In the second part, I discuss the relevant related work, such as Tanenhaus et al.'s "cue theory,"[6] which forms the theoretical basis for the present study. Additionally, I discuss the relevant applied NLP and machine learning literature. In the third part, I discuss the data, and in the fourth part I delve deeper into the methodology of a text classification-based approach to this question, with a discussion on text representation, modeling and issues of class imbalance. In the fifth part, I discuss the results and the limitations. Finally, I provide a conclusion, with discussions of future work.

1.2 United States Supreme Court's Case Selection Procedure

With such a low certiorari petition grant rate, it is perhaps unsurprising that there is a great deal of uncertainty surrounding what legal issues the Justices of the Supreme Court will deem deserving of review, or what makes a case "certworthy". To better understand what guides the Justices' decisions to grant or deny review, it is important to briefly examine the relevant procedure of the Supreme Court in this area.

The relevant procedures regarding certiorari petitions are based on both the written "Rules of the Supreme Court," which refer to internal rules, guidance, and procedures of the Supreme Court, as well as the unwritten customs of the Supreme Court. The "Rules of the Supreme Court" can be traced back to 1803 and have been constantly and consistently updated throughout the Court's existence, with the most recent version of the rules being adopted on December 5th, 2022, and made effective on January 1st, 2023. The most recent "Rules of the Supreme Court"[7] as well as the historical "Rules"[8] are openly available on the US Supreme Court's official website.

[6] Tanenhaus, Joseph, et al. "The Supreme Court's certiorari jurisdiction: Cue theory." *Judicial decision making* 111 (1963).
[7] Most recent "Rules of the Supreme Court" (Effective January 1, 2023) (PDF), available at: https://www.supremecourt.gov/filingandrules/2023RulesoftheCourt.pdf
[8] Historical "Rules of the Supreme Court", available at: https://www.supremecourt.gov/ctrules/scannedrules.aspx

1.2.1 The Rule of Four

The first relevant customary rule regarding certiorari petitions is known as the "Rule of Four", which establishes that a case will be heard by the US Supreme Court only if at least four out of nine Justices vote to grant review to a certiorari petition.[9] This non-majoritarian "Rule of Four"[10] can be compared to the more well-known "Rule of Five," which refers to the minimum number of Supreme Court Justices needed to establish a majority in a given case.

Being a customary rule, the "Rule of Four" procedure is not available in written form, and thus, the complete details of this procedure are esoteric and obscure by their very nature. In fact, somewhat surprisingly, even the history of the "Rule of Four" is unknown, with Leiman pointing out that, "research has yielded no evidence of [the Rule's] origin or early history."[11] Procedurally however, some peeks into the behind-the-scenes of this rule's operation have been made available in various cases,[12] most notably the *Rogers v. Missouri Pac. R. Co* (1957),[13] where Justices Frankfurter and Justice Harlan openly discuss their conflicting interpretations of what exactly the "Rule of Four" means. Nevertheless, for the present purposes, the simple and intuitive interpretation that the "Rule of Four" procedure necessarily entails that four out of nine Supreme Court Justices' votes are required to grant review for a given certiorari petition is sufficient.

1.2.2 Secret Voting without Explanation

Unlike written opinions of the US Supreme Court, where the observer knows exactly how each Justice voted and can read their reasoning and justifications for a decision, with regards to certiorari petitions, both the distribution of the votes and the information regarding how each Justice voted on a given petition are never revealed nor made available to the public. In other words, the certiorari review procedure is fundamentally secretive in nature. To further obscure the secretive nature of the certiorari selection process, traditionally, no opinion is written explaining the reasons why a certiorari petition was either granted or denied.

---

[9] Leiman, Joan Maisel. "The Rule of Four." *Colum. L. Rev.* 57 (1957): 975.
[10] Revesz, Richard L., and Pamela S. Karlan. "Nonmajority Rules and the Supreme Court." *U. Pa. L. Rev.* 136 (1987)
[11] Leiman, Joan Maisel. "The Rule of Four." *Colum. L. Rev.* 57 (1957): 975.
[12] Robbins, Ira P. "Justice by the Numbers: The Supreme Court and the Rule of Four-Or Is It Five." *Suffolk UL Rev.* 36 (2002)
[13] *Rogers v. Missouri Pacific R. Co.*, 352 U.S. 500, 77 S. Ct. 443, 1 L. Ed. 2d 493 (1957).

Consequently, a typical certiorari grant or denial contains only a single laconic sentence: "The petition for a writ of certiorari is granted/denied." Thus, Ulmer pertinently refers to this feature of the certiorari selection procedure as essentially being "secret decision-making without explanation."[14]

Nevertheless, some rare glimpses into the votes on certiorari petitions and the related legal reasoning are sometimes revealed in the form of "dissents from the denial of certiorari." An example of this can be seen in the recent case of *Dr A., et al. v Kathy Hochul, Governor of New York, et al.* 597 U. S. (2022)[15] concerning healthcare workers in New York who objected to vaccination requirements on religious grounds, but who were nevertheless subsequently "fired, forced to resign, lost admitting privileges, or been coerced into a vaccination."[16] In the "dissent from the denial of certiorari", Justice Thomas, who was joined by Justice Alito and Justice Gorsuch, wrote a short explanation as to why this particular certiorari petition should be granted review, noting the legal issues relating to the Free Exercise Clause under the First Amendment that the case brings up. However, such written "dissents from the denial of certiorari" exposing the Justices' reasoning and the associated votes are so incredibly rare that they can essentially be seen as *exceptio quae firmat regulam*, or "exceptions that prove the rule", the rule being that the certiorari petition voting is fundamentally secretive and that the reasoning behind certiorari votes is not revealed. It has been argued that the reason we do not see these types of "dissents from the denial of certiorari" more often stems from Supreme Court's need to show unity and the judiciary's fear that such dissents might expose which legal issues have a higher likelihood of gaining an audience with the Court.[17] Thus, the secretive and opaque nature of the certiorari selection process can make sense from a fairness perspective, that is, petitioners should not be able to game the system by filing certiorari petitions which are more likely to be granted review.

[14] Ulmer, S. Sidney. "The decision to grant certiorari as an indicator to decision" on the merits"." *Polity* 4.4 (1972)
[15] *Dr A., et al. v Kathy Hochul, Governor of New York, et al.* 597 U. S. (2022), Thomas J. dissenting from the denial of certiorari, available at: https://www.supremecourt.gov/opinions/21pdf/21-1143_3f14.pdf
[16] *Ibid*.
[17] Perry, Hersel W. Deciding to decide: agenda setting in the United States Supreme Court. Harvard University Press, 2009.

1.2.3 Certworthiness

Rule 10 of the "Rules of the Supreme Court" titled "Considerations Governing Review on Certiorari" governs what the Justices will take into account when voting to grant or deny certiorari petitions. It states that:

> "Review on a writ of certiorari is not a matter of right, but of judicial discretion.
> A petition for a writ of certiorari will be granted only for compelling reasons.
> The following, although neither controlling nor fully measuring the Court's discretion, indicate the character of the reasons the Court considers."[18]

Rule 10 proceeds to list out some examples of the "reasons the Court considers" when granting review, such as for instance a conflict between appellate courts or a conflict between state and appellate courts "on the same important matter."[19] Although this may seem helpful, it is important to reiterate that the rule itself states that these reasons are "neither controlling nor fully measuring the Court's discretion."[20] In other words, "certworthiness" seems to be entirely dependent on what individual Justices deem important, or as Perry puts it, an "important issue may be a highly technical issue of law, or it might be an issue of profound societal importance."[21] We can thus analogize the certiorari petition selection process to a black box algorithm where we have no conception of the modalities at play with regards to the case selection process. Again, the lack of transparency seems to be a fundamental feature of the system.

Some scholars have argued that this ambiguity and opaqueness of the certiorari selection, voting, and decision-making is indeed a deliberate choice because the Supreme Court purposefully tries to limit any possible inferences into what legal issues it deems "certworthy" or important enough to be granted review.[22] But taken together, these rules and customs paint a grim picture of a discretionary and secretive decision-making procedure, which, furthermore, raises questions of transparency, especially because of the highly consequential nature of certiorari petition decisions on America's society, law and politics.[23] These questions in turn have led some scholars to empirically analyze the Supreme Court's certiorari petition selection practice. As will be discussed later, a machine learning based text classification approach to certiorari petition prediction may offer a novel lens into the determinants of certiorari outcomes.

---

[18] "Rules of the Supreme Court" (Effective January 1, 2023) (PDF), available at: https://www.supremecourt.gov/filingandrules/2023RulesoftheCourt.pdf
[19] *Ibid.*
[20] *Ibid.*
[21] Perry, Hersel W. Deciding to decide: agenda setting in the United States Supreme Court. Harvard University Press, 2009.
[22] *Ibid.*
[23] Ulmer, S. Sidney. "The decision to grant certiorari as an indicator to decision" on the merits"." *Polity* 4.4 (1972)

1.3 Related Work and the Methodological Basis for Research

This section surveys previous work on the empirical research into certiorari petitions and the relevant applications of natural language processing methods, such as text classification, to questions in the legal domain. We will be adopting the insights from these works to propose a text classification-based methodological approach to the task of certiorari petition outcome prediction.

1.3.1 Cue Theory and Certiorari Outcome Prediction

The importance of the cases heard by the US Supreme Court, combined with non-transparent case selection procedure and the low certiorari grant rate has evoked interest in examining US Supreme Court's certiorari practice. The result is a large literature that deals with the problem of determining the predictors of certiorari outcomes, which is rooted primarily in Joseph Tanenhaus and associates' "cue theory" of Supreme Court certiorari jurisdiction.[24]

Tanenhaus et al. hypothesized that, given the immense workload of the Justices of the Supreme Court, "some method exists for separating the certiorari petitions requiring serious attention from those that are frivolous as to be unworthy of careful study."[25] Under the cue theory, the presence of "certain identifiable cues would warn a Justice that a petition deserved scrutiny."[26] In the words of Teger and Kosinski, "if cue theory is correct, only cases containing cues will be granted review; those without cues would be consigned to the footless appeals pile and not be voted on."[27]

---

[24] Tanenhaus, Joseph, et al. "The Supreme Court's certiorari jurisdiction: Cue theory." *Judicial decision making* 111 (1963).
See also: Teger, Stuart H., and Douglas Kosinski. "The cue theory of Supreme Court certiorari jurisdiction: A reconsideration." *The Journal of Politics* 42.3 (1980): 834-846.
See also: Ulmer, S. Sidney, William Hintze, and Louise Kirklosky. "The decision to grant or deny certiorari: Further consideration of cue theory." *Law & Society Review* 6.4 (1972): 637-643.
[25] Tanenhaus, Joseph, et al. "The Supreme Court's certiorari jurisdiction: Cue theory." *Judicial decision making* 111 (1963).
[26] *Ibid.*
[27] Teger, Stuart H., and Douglas Kosinski. "The cue theory of Supreme Court certiorari jurisdiction: A reconsideration." *The Journal of Politics* 42, no. 3 (1980): 834-846.

In their original work, Tanenhaus et al. examined four potential cues as predictors of review:

1) federal government as the party seeking review;
2) dissension, either among the judges of the court immediately below, or between two or more courts and agencies in a given case;
3) presence of a civil liberties issue;
4) presence of an economic issue.[28]

Tanenhaus et al. based their methodology on examining Supreme Courts' decisions and certiorari petitions for cues, but importantly for our present work, not on the text of the certiorari petitions themselves. This approach allowed them to observe which lower court decisions, and thus, which issues and which "cues" were deemed certworthy. They subsequently used chi square and phi statistics to examine the associations in their data, finding that only the fourth cue was not predictive of certiorari review.[29]

After the publication of Tanenhaus et al.'s study, the cue theory became the foundational theoretical framework for research on the predictors of certiorari outcomes. Accordingly, countless other cues have been argued to exist, for instance, Caldeira et al. propose that the filing of amicus curiae briefs in support of a certiorari petition gives the Justices "signals about the political, social, and economic significance of cases," which in turn might increase the likelihood that a certiorari petition will be granted review.[30] Recent work by Feldman and Kappner has argued that even names of entities involved in cases, such as for instance corporations or attorneys, can be predictive of the outcome of certiorari petitions.[31]

However, it is important to emphasize that throughout these studies the actual textual content of the certiorari petitions has not been fully utilized, which has two main drawbacks. First, the use of language arguably plays an important role in the legal domain,[32] and in the context of certiorari petitions, Perry states that some clerks attributed the consistent success of Solicitor-General's certiorari petitions to the fact the Solicitor

---

[28] Tanenhaus, Joseph, et al. "The Supreme Court's certiorari jurisdiction: Cue theory." *Judicial decision making* 111 (1963).
[29] *Ibid.*
[30] Caldeira, Gregory A., and John R. Wright. "Organized interests and agenda setting in the US Supreme Court." *American Political Science Review* 82.4 (1988): 1109-1127.
See also: Caldeira, Gregory A., and John R. Wright. "Amici curiae before the supreme court: who participates, when, and how much?." *The Journal of Politics* 52.3 (1990): 782-806.
[31] Feldman, Adam, and Alexander Kappner. "Finding certainty in cert: An empirical analysis of the factors involved in Supreme Court certiorari decisions from 2001-2015." Vill. L. Rev. 61 (2016): 795.
[32] Sim, Yanchuan, Bryan Routledge, and Noah Smith. "The utility of text: the case of Amicus briefs and the Supreme Court." In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 29, no. 1. 2015.

General "knows of all the catchwords, and they just know how to write them in a brief."[33] Second, from a practical standpoint, the Tanenhaus et al. methodology of reading numerous cases and labeling them for cues is undoubtedly an incredibly laborious process. But unfortunately, there are no viable substitutes to this approach, and according to Stern and Gressman practitioners have no alternative but to "rely as best as they can upon the history of grants and denials in the various types of cases."[34] Thus, for example, an attorney interested in whether their certiorari petition will be successful would likely have to read countless granted and denied certiorari petitions to even get a sense of the likelihood of their petition's success, but even after reading all the petitions and labeling them for their apparent cues, success is never guaranteed due to the discretionary and elusive nature of Supreme Court's certiorari petition selection procedure, making the costly work not worth the effort.

1.3.2 Supervised Machine Learning and Text Classification in the Legal Domain

Text classification, or text categorization, is one of the fundamental tasks in natural language processing.[35] Text classification is based on the supervised machine learning paradigm, namely it involves labeling data and subsequently building a classifier model that will be able to learn the features associated with each label or class.[36] The classifier model can then be used on unlabeled, or unseen data to make predictions as to their labels based on the features learned from the labeled data. In short, the classifier learns the features based on the labels determined by a human coder, thus directly utilizing the knowledge of the human coder. Fundamentally, according to Sebastiani, the supervised machine learning-based text classification paradigm can be divided into the following steps: text labeling, text representation and featurization, classifier construction and classifier evaluation.[37]

---

[33] Perry, Hersel W. Deciding to decide: agenda setting in the United States Supreme Court. Harvard University Press, 2009.

[34] Stern, Robert L., and Eugene Gressman. *Supreme Court practice: jurisdiction, procedure, arguing and briefing techniques, forms, statutes, rules for practice in the Supreme Court of the United States*. Bureau of national affairs, 1985.

[35] Cavnar, William B., and John M. Trenkle. "N-gram-based text categorization." *Proceedings of SDAIR-94, 3rd annual symposium on document analysis and information retrieval*. Vol. 161175. 1994.

See also: Joachims, Thorsten. "Text categorization with support vector machines: Learning with many relevant features." *European conference on machine learning*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998.

See also: Basu, Chumki, Haym Hirsh, and William Cohen. "Recommendation as classification: Using social and content-based information in recommendation." *Aaai/iaai*. 1998.

[36] Sebastiani, Fabrizio. "Machine learning in automated text categorization." *ACM computing surveys (CSUR)* 34.1 (2002): 1-47.

[37] *Ibid.*

Supervised machine learning-based text classification has proven to be an incredibly successful and powerful paradigm, both from the academic research and practical deployment perspectives. For instance, text classification methods have been applied in diverse areas of research, from analyzing sentiment in movie reviews,[38] to sarcasm detection on social media platforms like Twitter.[39] Supervised machine learning-based approaches have also been used with other types of data, such as for instance images, most famously, for the problem of handwritten digit recognition on the well-known MNIST database.[40] Fundamentally, the simple and highly interpretable supervised machine learning paradigm can be seen as being data type-agnostic and thus suitable for textual and other types of data.

While there has been an increase in the use of natural language processing methods to study legal questions,[41] text classification itself remains largely underutilized in the legal domain.[42] Nevertheless, there is some important recent work aiming to use text classification to explore legal questions. In particular, Aletras et al.'s work on the European Court of Human Rights (ECHR) decisions serves as the main guide to the methodological framework in the present work.[43] In their paper, Aletras et al. implement a binary text classification approach to predict and subsequently analyze the decisions of the ECHR with respect to specific Articles of the European Convention of Human Rights, using only text-based n-gram and topic features for representing the text of these decisions. Following Aletras et al, Sulea et al. implement a similar approach to the task of predicting the area of law in the decisions of the French Supreme Court.[44]

---

[38] Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. "Thumbs up? Sentiment classification using machine learning techniques." *arXiv preprint cs/0205070* (2002).
[39] Bamman, David, and Noah Smith. "Contextualized sarcasm detection on twitter." *proceedings of the international AAAI conference on web and social media*. Vol. 9. No. 1. 2015.
[40] Bottou, Léon, Corinna Cortes, John S. Denker, Harris Drucker, Isabelle Guyon, Larry D. Jackel, Yann LeCun et al. "Comparison of classifier methods: a case study in handwritten digit recognition." In *Proceedings of the 12th IAPR International Conference on Pattern Recognition, Vol. 3-Conference C: Signal Processing (Cat. No. 94CH3440-5)*, vol. 2, pp. 77-82. IEEE, 1994.
[41] Monroe, Burt L., Michael P. Colaresi, and Kevin M. Quinn. "Fightin' words: Lexical feature selection and evaluation for identifying the content of political conflict." *Political Analysis* 16, no. 4 (2008): 372-403;
See also: Danescu-Niculescu-Mizil, Cristian, Lillian Lee, Bo Pang, and Jon Kleinberg. "Echoes of power: Language effects and power differences in social interaction." In *Proceedings of the 21st international conference on World Wide Web*, pp. 699-708. 2012.
[42] Sulea, Octavia-Maria, Marcos Zampieri, Mihaela Vela, and Josef Van Genabith. "Predicting the law area and decisions of French Supreme Court cases." *arXiv preprint arXiv:1708.01681* (2017).
[43] Aletras, Nikolaos, Dimitrios Tsarapatsanis, Daniel Preoţiuc-Pietro, and Vasileios Lampos. "Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective." *PeerJ computer science* 2 (2016): e93.
[44] Sulea, Octavia-Maria, Marcos Zampieri, Mihaela Vela, and Josef Van Genabith. "Predicting the law area and decisions of French Supreme Court cases." *arXiv preprint arXiv:1708.01681* (2017).

In contrast to Aletras et al., the lack of rich textual information in the actual United States Supreme Court certiorari petition decisions necessitates a slight methodological reconsideration. As explained earlier, certiorari decisions are fundamentally opaque, and represent a procedure that has been aptly referred to as "secret decision-making without explanation."[45] Accordingly, grants or denials of certiorari petitions are only accompanied by a single uninformative sentence, namely that "The petition for a writ of certiorari is granted/denied." Therefore, our methodological approach, unlike Aletras et al., shifts the focus from the written judicial opinions to the textual content of the certiorari petitions themselves, which are written by petitioners who seek a review of their case by the Supreme Court. In this way, our method is more aptly comparable to supervised machine learning approaches to sentiment classification which involve the determination of whether a document is positive or negative.[46]

Therefore, to the best of our knowledge, no other study has applied supervised machine learning based text classification approaches to the task of certiorari petition outcome prediction.

1.4 Data

To implement a binary text classification-based approach to the task of certiorari outcome prediction we collected granted and denied petitions dealing with a specific Amendment of the US Constitution. We concentrate on examining a specific amendment following Aletras et al.'s methodology, which also focuses on specific ECHR Article violations.[47] Thus, our data consists of certiorari petitions relating to the Fourteenth Amendment because it plays a central role in civil rights litigations, and governs a broad range of issues such as equal protection, due process, and the application of the Bill of Rights to individual States through the doctrine of incorporation.[48] Previous research on cues, particularly the work of Teger and Kosinski, has also examined issues relating to the Fourteenth Amendment by decomposing the original Tanenhaus et al. "civil liberties" cue into its constituent subcategories.[49] In addition, from a data availability perspective, the Fourteenth Amendment provides a good data source for both granted and denied certiorari petitions. Analyzing outcomes of petitions dealing with Fourteenth Amendment

[45] Ulmer, S. Sidney. "The decision to grant certiorari as an indicator to decision" on the merits"." *Polity* 4.4 (1972)
[46] Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. "Thumbs up? Sentiment classification using machine learning techniques." arXiv preprint cs/0205070 (2002).
[47] Aletras, Nikolaos, Dimitrios Tsarapatsanis, Daniel Preoţiuc-Pietro, and Vasileios Lampos. "Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective." *PeerJ computer science* 2 (2016): e93.
[48] Amar, Akhil Reed. "The bill of rights and the fourteenth amendment." *Yale LJ* 101 (1991).
[49] Teger, Stuart H., and Douglas Kosinski. "The cue theory of Supreme Court certiorari jurisdiction: A reconsideration." *The Journal of Politics* 42, no. 3 (1980): 834-846.

will allow us to determine which issues under this Amendment are more likely to be granted or denied review. For future work we intend to explore other laws as well.

Our data source is LexisAdvance.[50] To ensure that our sample accurately represents Fourteenth Amendment certiorari petitions we download only those petitions that mention the Fourteenth Amendment at least five times. While this heuristic is not without limitations, we note that the results from the subsequent topic modeling analysis applied on our corpus gives us the prototypical Fourteenth Amendment topic descriptor words, which suggests that our data collection approach successfully captures certiorari petitions of interest. While the denied petitions were collected over a period from 2017 to 2006, we had to extend the time horizon for granted petitions to 1996 to address the severe lack of granted petitions under the Fourteenth Amendment and the related issue of class imbalance. Overall, our data collection approach yielded only 50 granted and 1012 denied petitions in total. It is worth pointing out that undoubtedly there are much more denied petitions in this period of time relating to the Fourteenth Amendment, but it is likely that not all of them have been digitized – data availability problems pose one of the main challenges for this research. When it comes to the textual content of the certiorari petitions, due to our interest in examining legal issues which are predictive of review across petitions we extracted the "Reasons for Granting" section of the certiorari petitions using regular expression matching.[51] The "Reasons for Granting" section is important because it provides the petitioner's explanation, legal reasoning and argumentation as to why a particular certiorari petition should be granted review.

[50] It is worth noting that LexisAdvance prohibits the scraping of data and bulk downloading, thus, data availability issues, particularly when it comes to denied certiorari petitions, pose a key challenge to text-based approaches to certiorari outcome prediction.

[51] Karttunen, Lauri, Jean-Pierre Chanod, Gregory Grefenstette, and Anne Schille. "Regular expressions for language engineering." *Natural Language Engineering* 2, no. 4 (1996): 305-328.

1.5 Methodology

The goal of the present work is to propose and apply a machine learning based methodology for analyzing and examining certiorari petition decisions based only on the text in said certiorari petitions. Using the text of petitions can greatly help legal professionals and academics with the task of petition outcome prediction and in the study of the issues predictive of certiorari review. Therefore, our hypothesis is that the textual content, and thus, the legal issues present in certiorari petitions, are predictive of review. To this end, the problem can be framed as a binary text classification task, which falls under a supervised machine learning paradigm.[52]

In the context of the supervised machine learning paradigm, we can frame the problem of certiorari petition outcome prediction as follows: the classifier f takes as input the textual representation of certiorari petitions, represented by the letter X. The goal of the classifier is to provide an output label, or a class, represented by the letter $\mathcal{Y}$, that is, either a petition has been granted, and thus labeled as 1, or the petition has been denied, and thus labeled as 0:

$$f\colon X \to \mathcal{Y}$$

Where:

f is the classifier,

X = textual representation of certiorari petitions, and

$\mathcal{Y}$ = {granted, denied}, or {1, 0}

1.5.1 Representation and Featurization

To implement the text classification-based approach, we must first transform the text of our documents into suitable numerical representations readable by the computer for subsequent analysis by machine learning algorithms, which is a process known as featurization.[53] To this end we first clean our corpus by lowercasing all the words, removing all non-alphabetical characters and stopwords,[54] and deploying the NLTK

[52] Joachims, Thorsten. "Text categorization with support vector machines: Learning with many relevant features." *European conference on machine learning*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998.
See generally: Vapnik, Vladimir. *The nature of statistical learning theory*. Springer science & business media, 1999.

[53] Joachims, Thorsten. "Text categorization with support vector machines: Learning with many relevant features." *European conference on machine learning*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998.

[54] Bird, Steven, Ewan Klein, and Edward Loper. *Natural language processing with Python: analyzing text with the natural language toolkit*. " O'Reilly Media, Inc.", 2009.

Wordnet Lemmatizer to lemmatize the words, which reduces the words in our corpus to their base forms known as "lemmas."[55] Lemmatization ensures that all forms of a word will refer to a single lemma. Thus, for example, after lemmatization, all instances of the word "petitions" will be stored as "petition" in our dataset. This ensures that the tokens, or individual instances of words – in our example, "petition" and "petitions" – are subsumed under a single lemma, or type, namely "petition", as it could be argued that they both refer to the same thing, with the only distinction being that "petitions" is a plural form of the word "petition."[56] Consistent reference to lemmas as opposed to tokens also reduces the dimensionality of the data, i.e. the number of features.[57]

It is important to point out some domain-specific information loss with regards to lemmatization. For instance, the word "damage" and "damages" have entirely different meanings in the legal domain. The token "damage" usually refers to "criminal damage", while conversely, the token "damages" usually refers to "monetary damages", which is a form of judicially awarded compensation. In other words, in the legal domain, the distinction between these two tokens "damage" and "damages" is not that of simple pluralization. Thus, lemmatizing "damages" can lead to important information loss as both "damage" and "damages" tokens will be subsumed under a single lemma "damage", even though both of these tokens refer to fundamentally distinct legal concepts. This problem of nuance with regards to representation of concepts is one of the limitations of natural language processing methods, especially in the context of specialized jargon unique to particular domains.[58]

Following Aletras et al., we use topic modeling to represent the petitions for our prediction task.[59] Topic modeling seems to be particularly well-suited for our task of finding common legal issues across petitions and examining which issues are predictive of review. The main advantage of using topic modeling lies in the significant reduction of the dimensionality of the data, in addition to the added interpretability of the output when compared to n-gram features.[60] The added interpretability of the topic modeling-based

---

[55] Miller, George A. "WordNet: a lexical database for English." *Communications of the ACM* 38, no. 11 (1995): 39-41.
[56] Peirce, Charles Santiago Sanders. "Prolegomena to an apology for pragmaticism." The Monist (1906): 492-546.
[57] Aggarwal, Charu C., and Charu C. Aggarwal. *Machine learning for text: An introduction*. Springer International Publishing, 2018.
[58] *Ibid.*
See generally: Jurafsky, Daniel, and James H. Martin. "Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition." 2023 ed.
[59] Aletras, Nikolaos, Dimitrios Tsarapatsanis, Daniel Preoţiuc-Pietro, and Vasileios Lampos. "Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective." *PeerJ computer science* 2 (2016): e93.
[60] *Ibid.*

classification output will allow one to examine which topics are associated with granted or denied classes and to what extent. Nevertheless, it is worth emphasizing that there is nothing that prevents using other text representation methods for certiorari outcome prediction.

For topic modeling we choose to use the non-Negative Matrix Factorization-based approach as the method for representing our petitions.[61] Non-Negative Matrix Factorization (NMF) is a dimensionality reduction method that approximately factorizes a non-negative matrix **A** into two factor matrices **W** and **H**, where all three matrices are subject to a non-negativity constraint.[62] This non-negativity constraint is particularly applicable to the text analytic domain, as document-term matrices do not contain negative elements.[63] In text analytic applications **W** represents the document-topic matrix, and **H** represents the topic-term matrix, which, when multiplied together, approximately reconstruct the matrix **A**:[64]

$$\mathbf{A} \approx \mathbf{W} \times \mathbf{H}$$

Where:

A is the original document-term matrix representation of the corpus;
W is the smaller document-topic matrix;
H is the smaller topic-term matrix.

NMF is advantageous for analyzing niche and non-mainstream content[65] which is particularly suitable in the context of our certiorari petition dataset which contains considerable domain-specific legal jargon. A further benefit of using NMF is its good handling of polysemy and synonymy combined with high interpretability of the resulting topics and descriptor words.[66] Nevertheless, it is worth pointing out there are other topic

---

See also: Quinn, Kevin M., Burt L. Monroe, Michael Colaresi, Michael H. Crespin, and Dragomir R. Radev. "How to analyze political attention with minimal assumptions and costs." *American Journal of Political Science* 54, no. 1 (2010): 209-228.

[61] Lee, Daniel D., and H. Sebastian Seung. "Learning the parts of objects by non-negative matrix factorization." *Nature* 401, no. 6755 (1999): 788-791.

[62] *Ibid.*

[63] Aggarwal, Charu C., and Charu C. Aggarwal. *Machine learning for text: An introduction*. Springer International Publishing, 2018.

[64] *Ibid.*

[65] O'callaghan, Derek, Derek Greene, Joe Carthy, and Pádraig Cunningham. "An analysis of the coherence of descriptors in topic modeling." *Expert Systems with Applications* 42, no. 13 (2015): 5645-5657.

[66] Aggarwal, Charu C., and Charu C. Aggarwal. *Machine learning for text: An introduction*. Springer International Publishing, 2018.

See also: Brunet, Jean-Philippe, Pablo Tamayo, Todd R. Golub, and Jill P. Mesirov. "Metagenes and molecular pattern discovery using matrix factorization." *Proceedings of the national academy of sciences* 101, no. 12 (2004): 4164-4169.

modeling methods, such as Latent Dirichlet Allocation (LDA)[67] and others[68] which rely on different mathematical modeling approaches and assumptions. Due to the applied nature of the present work, the intricate subject of the mathematical distinctions between topic modeling methods lies outside of the scope of this paper.

We implement Non-Negative Matrix Factorization with Non-Negative Double Singular Value Decomposition (NNDSVD) initialization[69] on a Term Frequency-Inverse Document Frequency (TF-IDF) weighted matrix of unigrams[70] that appear more than 2 times, with the optimal number of topics set to $k = 53$. To find the k optimal number of topics, we implemented the Topic Coherence-Word2Vec (TC-W2V) approach of calculating the mean pairwise cosine similarity of top 20 descriptor terms for each topic in the NMF models built from $k = 3$ to $k = 100$.[71] The pairwise cosine similarities in turn are obtained from a 100-dimensional word2vec skip-gram word embedding model built on the entire corpus.[72]

1.5.2 Classification Models

Having discussed the data and the representation methods we will be utilizing, we proceed to discuss the models which will be used for binary classification of granted and denied certiorari petitions.

The first model we implement is a Linear Kernel Support Vector Machine (Linear SVM).[73] Support Vector Machines (SVM) is a learning method that has been shown to be highly effective in various machine learning classification tasks, including text

---

[67] Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent dirichlet allocation." *Journal of machine Learning research* 3, no. Jan (2003): 993-1022.
[68] Quinn, Kevin M., Burt L. Monroe, Michael Colaresi, Michael H. Crespin, and Dragomir R. Radev. "How to analyze political attention with minimal assumptions and costs." *American Journal of Political Science* 54, no. 1 (2010): 209-228.
[69] Boutsidis, Christos, and Efstratios Gallopoulos. "SVD based initialization: A head start for nonnegative matrix factorization." *Pattern recognition* 41, no. 4 (2008): 1350-1362.
[70]Luhn, Hans Peter. "A statistical approach to mechanized encoding and searching of literary information." *IBM Journal of research and development* 1, no. 4 (1957): 309-317.
See also: Salton, Gerard, and Christopher Buckley. "Term-weighting approaches in automatic text retrieval." *Information processing & management* 24, no. 5 (1988): 513-523.
[71] O'callaghan, Derek, Derek Greene, Joe Carthy, and Pádraig Cunningham. "An analysis of the coherence of descriptors in topic modeling." *Expert Systems with Applications* 42, no. 13 (2015): 5645-5657.
[72] Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. "Distributed representations of words and phrases and their compositionality." *Advances in neural information processing systems* 26 (2013).
[73] Cortes, Corinna, and Vladimir Vapnik. "Support-vector networks." *Machine learning* 20 (1995): 273-297.

classification.[74] We implement a Linear Kernel SVM as it allows us to rank our features[75] – in our case these features are NMF topics which can be interpreted as representing legal issues – and their associations with certiorari grants or denials, which has the benefit of allowing us to make some qualitative statements about US Supreme Court's certiorari petition selection patterns and decisions.[76] The second model we use is Extreme Gradient Boosting (XGBoost), which is a popular tree boosting learning method known for its good performance on various machine learning tasks.[77] For future work on this research project, we will also be implementing the open-source Pycaret machine learning library[78] which allows the researcher to train multiple classification models simultaneously and choose the best performing model for the task. As noted earlier, due to the applied nature of our present work, the choice of classification models and the underlying mathematical and statistical details lie outside the scope of our study.

1.5.3 Class Imbalance

With 1012 denied, and 50 granted certiorari petitions, our data is severely imbalanced, and a number of problems arise when dealing with data imbalance issues of this kind.[79] The first major problem is that it becomes difficult to accurately measure the predictive performance of our classification models. For instance, Lopez et al. have demonstrated that stratified k-fold cross validation, which is one of the traditional model selection approaches used in machine learning, becomes problematic in imbalanced data settings due to the fact that "a single misclassified example of the positive class can

[74] Joachims, Thorsten. "Text categorization with support vector machines: Learning with many relevant features." *European conference on machine learning*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998.
See also: Vapnik, Vladimir. *The nature of statistical learning theory*. Springer science & business media, 1999.
See also: Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. "Thumbs up? Sentiment classification using machine learning techniques." *arXiv preprint cs/0205070* (2002).
[75] Chang, Yin-Wen, and Chih-Jen Lin. "Feature ranking using linear SVM." In *Causation and prediction challenge*, pp. 53-64. PMLR, 2008.
[76] Aletras, Nikolaos, Dimitrios Tsarapatsanis, Daniel Preoţiuc-Pietro, and Vasileios Lampos. "Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective." *PeerJ computer science* 2 (2016): e93.
[77] Chen, Tianqi, and Carlos Guestrin. "Xgboost: A scalable tree boosting system." In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785-794. 2016.
[78] https://pycaret.org/
[79] He, Haibo, and Yunqian Ma, eds. "Imbalanced learning: foundations, algorithms, and applications." (2013).
See also: Fernández, Alberto, Salvador García, Mikel Galar, Ronaldo C. Prati, Bartosz Krawczyk, and Francisco Herrera. Learning from imbalanced data sets. Vol. 10. Cham: Springer, 2018.

create a significant drop in performance."[80] The class imbalance problem is also exacerbated by the prevalence of the so-called "dataset shift" problems, which occur when "the joint distribution of inputs and outputs differs between training and test stage."[81] In other words, the training distribution can be fundamentally different to the test distribution, both of which are assumed to be identical in traditional machine learning.[82] In the case of our present research, the fact that there are very few granted petitions and many denied petitions, both of which are collected over a number of years, engenders this problem due to random selection of training and test samples, and thus, can give inaccurate evaluations with regards to model selection.[83] Therefore, to evaluate our models we adopt the Monte Carlo-cross validation approach (MCCV) because it offers an asymptotically consistent method for model selection, which is, nonetheless, computationally expensive.[84] Under MCCV, the data is randomly partitioned into training and testing sets N number of times and the model is evaluated on each of these iterations with the relevant metrics averaged over N.[85] In our case, the test set at each iteration is 20% of the data and we run MCCV 10^3 times for each model. As stated earlier, MCCV is computationally expensive, so we are unable to run MCCV for more than 10^3 times due to working on a single computer without a GPU, however, for future work, we intend to utilize GPU clusters.

---

[80] López, Victoria, Alberto Fernández, Salvador García, Vasile Palade, and Francisco Herrera. "An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics." *Information sciences* 250 (2013): 113-141.

[81] Quinonero-Candela, Joaquin, Masashi Sugiyama, Anton Schwaighofer, and Neil D. Lawrence, eds. *Dataset shift in machine learning*. Mit Press, 2008.

[82] *Ibid.*

[83] López, Victoria, Alberto Fernández, and Francisco Herrera. "On the importance of the validation technique for classification with imbalanced datasets: Addressing covariate shift when data is skewed." *Information Sciences* 257 (2014): 1-13.

[84] Shan, Guogen. "Monte Carlo cross-validation for a study with binary outcome and limited sample size." *BMC Medical Informatics and Decision Making* 22, no. 1 (2022): 1-15.
See also: A. Ramezan, Christopher, Timothy A. Warner, and Aaron E. Maxwell. "Evaluation of sampling and cross-validation tuning strategies for regional-scale machine learning classification." *Remote Sensing* 11, no. 2 (2019): 185.
See also: Khalilia, Mohammed, Sounak Chakraborty, and Mihail Popescu. "Predicting disease risks from highly imbalanced data using random forest." *BMC medical informatics and decision making* 11 (2011): 1-13.
See also: Xu, Qing-Song, and Yi-Zeng Liang. "Monte Carlo cross validation." *Chemometrics and Intelligent Laboratory Systems* 56, no. 1 (2001): 1-11.
See also: Picard, Richard R., and R. Dennis Cook. "Cross-Validation of Regression Models." *Journal of the American Statistical Association* 79, no. 387 (1984): 575–83. https://doi.org/10.2307/2288403.
See also: Burman, Prabir. "A comparative study of ordinary cross-validation, v-fold cross-validation and the repeated learning-testing methods." *Biometrika* 76, no. 3 (1989): 503-514.
See also: Shao, Jun. "Linear model selection by cross-validation." *Journal of the American statistical Association* 88, no. 422 (1993): 486-494.

[85] Burman, Prabir. "A comparative study of ordinary cross-validation, v-fold cross-validation and the repeated learning-testing methods." *Biometrika* 76, no. 3 (1989): 503-514.

Another feature of machine learning from imbalanced datasets is that accuracy alone, which is one of the "most widely used basic measures of classifier performance,"[86] no longer provides adequate information on the classifier's performance.[87] Definitionally, accuracy simply "asks what percentage of all the observations our system labeled correctly,"[88] and is mathematically defined as the ratio of those observations correctly labeled as "true positive" and "true negative" divided by the total number of observations.[89] The problem of using accuracy for imbalanced data problems, such as the present certiorari petition outcome prediction task, lies precisely in the overabundant nature of the denied class. Thus, for instance, a "dummy classifier," which is "a non-learning model that makes predictions following a simple set of rules, for example, always predicting the most frequent class label"[90] if applied to the present task would have incredibly high accuracy simply due to the large amount of denied petitions that the certiorari process generates. Thus, as Jurafsky and Martin state, "accuracy doesn't work well when the classes are unbalanced and is not a good metric when the goal is to discover something that is rare, or at least not completely balanced in frequency, which is a very common situation in the world."[91]

Therefore, in the context of the certiorari outcome prediction problem, evaluation of classifier performance based primarily on recall and not on accuracy thus seems more appropriate.[92] The recall metric, which is also known as sensitivity,[93] "measures the percentage of items actually present in the input that were correctly identified"[94] and is defined as the ratio of observations labeled as "true positive" over the sum of "true positives" and "false negatives."[95] In the aforementioned example of a "dummy classifier" which predicts every petition as denied, its recall would be zero since there

---

[86] Saito, Takaya, and Marc Rehmsmeier. "The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets." *PloS one* 10, no. 3 (2015): e0118432.
[87] He, Haibo, and Edwardo A. Garcia. "Learning from imbalanced data." *IEEE Transactions on knowledge and data engineering* 21, no. 9 (2009): 1263-1284.
[88] Jurafsky, Daniel, and James H. Martin. "Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition." 2023 ed.
[89] *Ibid.*
[90] van de Bijl, Etienne, Jan Klein, Joris Pries, Sandjai Bhulai, Mark Hoogendoorn, and Rob van der Mei. "The Dutch Draw: constructing a universal baseline for binary prediction models." *arXiv preprint arXiv:2203.13084* (2022).
[91] Jurafsky, Daniel, and James H. Martin. "Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition." 2023 ed.
[92] He, Haibo, and Edwardo A. Garcia. "Learning from imbalanced data." *IEEE Transactions on knowledge and data engineering* 21, no. 9 (2009): 1263-1284.
[93] Swets, John A. "Measuring the accuracy of diagnostic systems." *Science* 240, no. 4857 (1988): 1285-1293.
[94] Jurafsky, Daniel, and James H. Martin. "Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition." 2023 ed.
[95] *Ibid.*

would be no true positives detected by the classifier.[96] Therefore, recall, which can also be aptly called the "completeness or effectiveness of retrieval,"[97] evaluates model performance by "indicating that the algorithm is performing well in classifying true positives and minimizes false negatives."[98] For our present task, recall is thus important due to the high cost associated with misclassifying the rare true positive observations, or granted certiorari petitions, in our sample. Another important metric in imbalanced data settings is precision, or "purity of retrieval"[99] which is defined as "the percentage of the items that the system detected that are in fact positive,"[100] and it is mathematically defined as the ratio of observations labeled as "true positive" over the sum of "true positives" and "false positives."[101] Intuitively, precision is "appropriate when [the minimization of] false positives is the focus,"[102] whereas recall is conversely "appropriate when [the minimization of] false negatives is the focus."[103] Furthermore, in the words of He ad Ma, "in imbalanced datasets, the goal is to improve recall without hurting precision. These goals, however, are often conflicting, since in order to increase the true positive for the minority class the number of false positives is also often increased, resulting in reduced precision."[104] In the context of our problem, since the misclassification of denied petitions as granted is not as critical as the misclassification of granted petitions as denied, precision, although relevant to evaluation, is not as important as recall. In other words, since we care more about minimizing false negatives we focus generally on recall for evaluating classifier performance.[105] Our emphasis on recall rather than precision also makes intuitive and practical sense, since an attorney or a researcher can simply examine the petitions that the classifier predicted will be granted and subsequently rank them based on their respective probabilities, which is a viable alternative to reading all the certiorari petitions in the corpus. This is consistent with the traditional heuristics developed in the document retrieval literature, where "the user can narrow a list of retrieved items by actually reading them."[106]

---

[96] *Ibid.*
[97] Buckland, Michael, and Fredric Gey. "The relationship between recall and precision." *Journal of the American society for information science* 45, no. 1 (1994): 12-19.
[98] Johri, Prashant, Mario José Diván, Ruqaiya Khanam, Marcelo Marciszack, and Adrián Will, eds. *Trends and advancements of image processing and its applications*. Springer International Publishing, 2022.
[99] *Ibid.*
[100] Jurafsky, Daniel, and James H. Martin. "Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition." 2023 ed.
[101] *Ibid.*
[102] Brownlee, Jason. *Imbalanced classification with Python: better metrics, balance skewed classes, cost-sensitive learning*. 2020.
[103] *Ibid.*
[104] He, Haibo, and Yunqian Ma, eds. "Imbalanced learning: foundations, algorithms, and applications." (2013).
[105] Brownlee, Jason. *Imbalanced classification with Python: better metrics, balance skewed classes, cost-sensitive learning*. 2020.
[106] Basu, Chumki, Haym Hirsh, and William Cohen. "Recommendation as classification: Using social and content-based information in recommendation." *Aaai/iaai*. 1998.

Some other mitigation strategies for imbalanced learning problems include approaches which generate synthetic data, such as synthetic minority over-sampling technique (SMOTE),[107] which has become the "standard benchmark for learning from imbalanced data."[108] SMOTE algorithm is a simple algorithm that works by creating synthetic data via "interpolation between several minority class instances that are within a defined neighborhood."[109] In our case that would essentially generate synthetic topic-based representations via the SMOTE algorithm's exploration of the relationship between document representations in the minority class feature space. However, it is worth pointing out that after trying out this rather simple but effective method, we unfortunately did not gain any significant improvements in classifier performance. For future work other minority oversampling techniques suggested in the literature will also be explored.[110]

1.6 Results and Discussion

The average Monte Carlo Cross-Validation (MCCV) classification accuracy, recall, and precision scores for both models trained and tested on topic features after 10^3 simulations are shown below in Table 1:

Classifier	Recall	Precision	Accuracy
Linear SVM	56.6	11.3	77.1
XGBoost	72.2	7.1	53.5

Table 1: Average Monte Carlo Cross-Validation (MCCV) scores over 10^3 simulations in percent

We can see that the Linear SVM model does better on accuracy, but worse on recall, whereas XGboost is able to achieve the opposite results.[111] Given our interest in the

[107] Chawla, Nitesh V., Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. "SMOTE: synthetic minority over-sampling technique." Journal of artificial intelligence research 16 (2002): 321-357.

[108] Fernández, Alberto, Salvador Garcia, Francisco Herrera, and Nitesh V. Chawla. "SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary." *Journal of artificial intelligence research* 61 (2018): 863-905.

[109] *Ibid.*

[110] He, Haibo, Yang Bai, Edwardo A. Garcia, and Shutao Li. "ADASYN: Adaptive synthetic sampling approach for imbalanced learning." In *2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence)*, pp. 1322-1328. Ieee, 2008.
See also: Brandt, Jakob, and Emil Lanzén. "A comparative review of SMOTE and ADASYN in imbalanced data classification." (2021).

[111] On the relationship between recall and precision see generally: Buckland, Michael, and Fredric Gey. "The relationship between recall and precision." *Journal of the American society for information science* 45, no. 1 (1994): 12-19.

classifier's recall, as discussed in section 1.5.3 on "Class Imbalance," XGBoost seems to be the better model for this particular classification problem and with our particular certiorari petition data. Furthermore, when it comes to both recall and accuracy, both of these models' results surpass the random choice baseline of 50%. Thus, the results above provide suggestive evidence confirming the hypothesis that text-based features in certiorari petitions, particularly topic-based features, do contain predictive power when it comes to Fourteenth Amendment certiorari petitions.

To qualitatively explore our results, we extract the model weights from the Linear SVM model, as shown in Table 2 below. Linear SVM weights rank the features necessary for optimal classification which can be indicative of membership in a granted or denied petition class.[112] As we can see, the Linear SVM model weights indicate that issues of discrimination – as represented by topic descriptor words in Topic 47 – would be granted review under the Fourteenth Amendment. This makes intuitive sense because the Fourteenth Amendment specifically deals with issues of discrimination, equal protection, and incorporation of the Bill of Rights against States. On the other hand, certiorari petitions containing descriptor words in Topic 37 – which seems to be dealing with cases on President Obama's citizenship – are more likely to be denied review under the Fourteenth Amendment. This also makes intuitive sense likely because of the frivolous nature of this issue. Furthermore, the topic most predictive of certiorari grants is Topic 22, which, based on the topic descriptor words, seems to refer to issues relating to juvenile sentencing and death penalty. Interestingly, Perry's seminal work has shown that Supreme Court Justices seem to be especially sensitive to certiorari petitions dealing with death penalties.[113] For instance, a Supreme Court clerk interviewed by Perry pointed out that "death penalty cases are treated very, very seriously," with another clerk stating that "all death penalties are discussed [by the Justices] at conference."[114] Furthermore, in the context of juvenile death penalties, Lain notes that "in 2002, four Justices dissented from the Court's denial of certiorari in the original *Stanford* case, writing, "We should put an end to this shameful practice."[115] Thus, it is clear that issues of death penalty, especially juvenile death penalty, will likely be predictive of review, and therefore, it is reassuring that our proposed method, despite some limitations, does seem to confirm existing insights into Supreme Court's certiorari selection practice.

---

[112] Guyon, Isabelle, Jason Weston, Stephen Barnhill, and Vladimir Vapnik. "Gene selection for cancer classification using support vector machines." *Machine learning* 46 (2002): 389-422. Aletras, Nikolaos, Dimitrios Tsarapatsanis, Daniel Preoţiuc-Pietro, and Vasileios Lampos. "Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective." *PeerJ computer science* 2 (2016): e93.

[113] Perry, Hersel W. *Deciding to decide: agenda setting in the United States Supreme Court*. Harvard University Press, 2009.

[114] *Ibid*. p 95.

[115] Lain, Corinna Barrett. "Deciding death." *Duke LJ* 57 (2007): 1. p 75
See also: *In re Stanford*, 537 U.S. 968, 972 (2002), Stevens, J., dissenting from denial of certiorari

While our inference is preliminary, the proposed methodology seems to offer results that we think are highly interpretable. Thus, this approach can indeed be valuable to legal researchers and practitioners as it would allow them to examine the Supreme Court's petition decisions without the need to closely read granted and denied petitions. Additionally, due to the data agnostic nature of text classification approaches, this method can also be applied to other types of legal motions and petitions in other courts and indeed other countries and jurisdictions, including outside the United States. For instance, UK's Supreme Court has a similar case selection procedure known as the "permission to appeal" process,[116] which can also be investigated using the proposed text classification-based methodology.

Topic No	Topic Name	Topic Descriptor Words	SVM Feature Weights
22	Juvenile sentencing and death penalty	juvenile, sentence, adult, apprendi, death, parole	6.24238713
39	Issues relating to criminal procedure	jeopardy, double, criminal, conviction, offense, acquittal,	5.13999871
47	Discrimination, Title VII of the Civil Rights Act 1964	title, congress, discrimination, vii, ada, adea	4.71371722
…	…	…	…
29	Injuries due to malpractice	statute, limitation, period, injury, claim, malpractice	–8.01075608
50	Drugs used in administering the death penalty	execution, protocol, arthur, inmate, drug, midazolam	–10.23758512
37	Cases questioning President Obama's citizenship	born, citizen, natural, president, obama, candidate	–12.85090526

Table 2: Topics predictive of Fourteenth Amendment certiorari grants and denials. In total there are $k = 53$ topics, and the three most predictive topics for the denied and granted classes are shown. The top six topic descriptor words for each topic are also shown. Topic names are inferred after interpreting the topic descriptor words. Positive feature weights in the "SVM Feature Weight" column indicate the most predictive topics for certiorari petition grants – i.e., the legal issues in certiorari petitions most likely to succeed and be granted review by the Supreme Court. Conversely, negative feature weights in the "SVM Feature Weight" column indicate the most predictive topics for certiorari petition denials – i.e., the legal issues in certiorari petitions most likely to be denied review by the Supreme Court.

[116] Hanretty, Chris. *A court of specialists: Judicial behavior on the UK Supreme Court*. Oxford University Press, USA, 2020.

As we discussed earlier, there are, however, a number of limitations. First, the intrinsic class imbalance of the underlying granted and denied certiorari petition generation process poses a great obstacle from both the data collection and model development perspectives.[117] For future work, we intend to expand our dataset by utilizing certiorari petition data regardless of the legal issue in question, thus departing from Aletras' et al.'s approach of focusing on specific legal articles. Furthermore, the incorporation of other "cues" as features, such as for instance the number or textual content of amicus curiae briefs,[118] indications of circuit splits,[119] political leanings at the court,[120] or other data sources,[121] etc., seem to be the clear next steps in our research agenda.

It is worth also pointing out that a fundamental problem of the certiorari outcome prediction task is the similarity in the language used across granted and denied petitions. Indeed, this is consistent with literature on Supreme Court certiorari practice. For instance, Perry points out that many times, Justices may consistently deny certiorari because they wait for "good vehicle" cases that contain "good facts."[122] In other words, justices might deny a perfectly good case because it's not the "perfect case" with the "perfect facts" for their agenda, whatever it may be. Thus, for example, a case that was denied review three years ago might be granted review today, simply because the issue becomes important only now. However, if the data includes both of these identical cases it might pose a problem for classification. Overall, this shows that certiorari petition decisions are extremely complex, and likely as polarized as the substantive Supreme Court decisions, but despite that, the textual content of certiorari petitions can indeed offer a new dimension for research and analysis of certiorari outcomes.

Furthermore, even though the focus of this paper has been on examining the utility and value of applying a natural language processing based text classification paradigm to the task of certiorari outcome prediction, the results, nevertheless, seem to confirm the intuition that Supreme Court Justices' decisions to grant or deny certiorari review do not necessarily depend solely on the text of the petitions.[123] This in turn, confirms the

---

[117] He, Haibo, and Edwardo A. Garcia. "Learning from imbalanced data." *IEEE Transactions on knowledge and data engineering* 21, no. 9 (2009): 1263-1284.
[118] Caldeira, Gregory A., and John R. Wright. "Amici curiae before the supreme court: who participates, when, and how much?." *The Journal of Politics* 52.3 (1990): 782-806.
[119] Stephenson, Ryan. "Federal Circuit Cases Selection at the Supreme Court: An Empirical Analysis." *Geo. LJ* 102 (2013): 271.
[120] Martin, Andrew D., and Kevin M. Quinn. "Dynamic ideal point estimation via Markov chain Monte Carlo for the US Supreme Court, 1953–1999." *Political analysis* 10, no. 2 (2002): 134-153.
[121] I am currently working with two other researchers on using the proposed approach to certiorari outcome prediction but with more data.
[122] Perry, Hersel W. Deciding to decide: agenda setting in the United States Supreme Court. Harvard University Press, 2009.
[123] *Ibid.*

complex nature of the underlying decision-making space in that the actual reasons for granting or denying a certiorari petition might be entirely independent of the text of the certiorari petitions, for instance, perhaps because of political or societal reasons, or other, unknown reasons for granting review could be at play, necessitating a healthy skeptical approach to the utility of text analytic methods for research into certiorari petition outcome prediction. Intuitively, if indeed it was the case that political or social considerations played no role in Supreme Court Justices' decision-making, then Supreme Court Justice nominations and confirmation would not be so politically contentious.[124] It would thus be beneficial to include other features into the prediction task, such as the political leanings of the Justices of the Court,[125] or the political party of the president during the Court's tenure, etc. Overall, without an increase in transparency of the certiorari petition selection process, it will be difficult to state conclusively what the actual underlying reasons for the Court's certiorari grant and denial decisions really are.

1.7 Conclusion

According to Hersel Perry, the "prediction of certiorari is so difficult in part because of the idiosyncratic interests in particular areas that lead some Justices, and not others, to see certain issues as important."[126] In this present research, we tried to find out which legal issues under the Fourteenth Amendment are deemed important by Supreme Court Justices. To this end, we adopted natural language processing methods such as text classification and topic modeling to the task of certiorari outcome prediction.

To discover which legal issues in certiorari petitions are predictive of review, we thus proposed adopting a binary text classification paradigm with topic-based features. We began by exploring previous related research on the determinants of the Supreme Court's certiorari petition decisions, notably Tanenhaus et al.'s "cue theory" of certiorari petition selection.[127] Furthermore, we explored the use of text-based topic features to represent legal issues for the prediction of certiorari petition grants or denials under the Fourteenth Amendment. Our work shows that the proposed text classification paradigm can indeed be successfully applied to the problem of certiorari outcome prediction. Therefore, since under the "cue theory" certain cues are argued to be predictive of review, our results not only support Tanenhaus' et al's "cue theory," but furthermore, the proposed methodology

[124] Kastellec, Jonathan P., Jeffrey R. Lax, Michael Malecki, and Justin H. Phillips. "Polarizing the electoral connection: partisan representation in Supreme Court confirmation politics." *The journal of politics* 77, no. 3 (2015): 787-804.
[125] Martin, Andrew D., and Kevin M. Quinn. "Dynamic ideal point estimation via Markov chain Monte Carlo for the US Supreme Court, 1953–1999." *Political analysis* 10, no. 2 (2002): 134-153.
[126] *Ibid.*
[127] Tanenhaus, Joseph, et al. "The Supreme Court's certiorari jurisdiction: Cue theory." *Judicial decision making* 111 (1963).

enables a text-based approach to the problem which allows for a more granular qualitative examination of the cues and the legal issues present in the certiorari petitions.

Future work can proceed in multiple directions. In addition to investigating other textual representation techniques, such as n-gram based representations,[128] word embeddings,[129] or the more advanced BERT-based representations of documents,[130] we also intend to incorporate other possible relevant features, such as for instance the political leanings of judges on the Supreme Court,[131] among other features which we discussed in the sections above. Secondly, the underlying class imbalance problem necessitates collection of more data or a possible adoption of synthetic data approaches to improve classifier performance. Finally, we also intend to study the historical patterns of "dissents from the denial of certiorari", including the textual patterns contained in these dissents to investigate whether these "dissents from the denial of certiorari" are increasing or declining in frequency over time, particularly in the context of the growing questions surrounding the polarization at the United States Supreme Court.[132]

As discussed earlier, our proposed approach suffers from a number of limitations, such as the aforementioned class imbalance of the underlying certiorari petition data generating process. The data problems reflect the complexity of the feature space, as is the case when Justices wait for the "perfect case."[133] Furthermore, there are important omitted variable issues, reflected in the fact that the reasons for granting or denying a certiorari petition could be completely outside the scope of the text of a petition, such as for instance, when the political aspects in the country determine Supreme Court's Justices' decisions, or when Justices seek to decide on societal issues that are deemed important.

By proposing and implementing a supervised machine learning-based binary text classification approach, we were successfully able to get a glimpse into what issues under

---

[128] Bamman, David, Jacob Eisenstein, and Tyler Schnoebelen. "Gender identity and lexical variation in social media." *Journal of Sociolinguistics* 18, no. 2 (2014): 135-160.
See also: Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. "Thumbs up? Sentiment classification using machine learning techniques." *arXiv preprint cs/0205070* (2002).
[129] Lilleberg, Joseph, Yun Zhu, and Yanqing Zhang. "Support vector machines and word2vec for text classification with semantic features." In 2015 IEEE 14th International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC), pp. 136-140. IEEE, 2015.
[130] Lin, Yuxiao, Yuxian Meng, Xiaofei Sun, Qinghong Han, Kun Kuang, Jiwei Li, and Fei Wu. "Bertgcn: Transductive text classification by combining gcn and bert." arXiv preprint arXiv:2105.05727 (2021).
[131] Martin, Andrew D., and Kevin M. Quinn. "Dynamic ideal point estimation via Markov chain Monte Carlo for the US Supreme Court, 1953–1999." Political analysis 10.2 (2002): 134-153.
[132] *Ibid.*
See also: Devins, Neal, and Lawrence Baum. "Split definitive: How party polarization turned the Supreme Court into a partisan court." The Supreme Court Review 2016, no. 1 (2017): 301-365.
[133] Perry, Hersel W. Deciding to decide: agenda setting in the United States Supreme Court. Harvard University Press, 2009.

the Fourteenth Amendment are deemed important enough for review, without the need to closely read the thousands of granted and denied certiorari petitions. This is an important benefit of the proposed approach, as it would allow the researcher and the practitioner to quickly sift through thousands of certiorari petitions and gather meaningful insights which will then be supported by further investigation and close reading. Another benefit of the proposed text classification approach is that it can be extended to other areas of law and other types of motions and petitions, including petitions in other countries and jurisdictions. Larger datasets and incorporation of other features into the prediction task, such as for instance other "cues" or data on political leanings of Justices, amicus curiae, etc., among others, are some of the next steps in this research.

1.8 References

Aggarwal, Charu C., and Charu C. Aggarwal. Machine learning for text: An introduction. Springer International Publishing, 2018.

Aletras, Nikolaos, Dimitrios Tsarapatsanis, Daniel Preoţiuc-Pietro, and Vasileios Lampos. “Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective.” PeerJ computer science 2 (2016): e93.

Amar, Akhil Reed. "The bill of rights and the fourteenth amendment." Yale LJ 101 (1991).

Bamman, David, and Noah Smith. "Contextualized sarcasm detection on twitter." proceedings of the international AAAI conference on web and social media. Vol. 9. No. 1. 2015.

Bamman, David, Jacob Eisenstein, and Tyler Schnoebelen. "Gender identity and lexical variation in social media." Journal of Sociolinguistics 18, no. 2 (2014): 135-160.

Basu, Chumki, Haym Hirsh, and William Cohen. "Recommendation as classification: Using social and content-based information in recommendation." Aaai/iaai. 1998.

Bird, Steven, Ewan Klein, and Edward Loper. Natural language processing with Python: analyzing text with the natural language toolkit. " O'Reilly Media, Inc.", 2009.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent dirichlet allocation." Journal of machine Learning research 3, no. Jan (2003): 993-1022.

Bottou, Léon, Corinna Cortes, John S. Denker, Harris Drucker, Isabelle Guyon, Larry D. Jackel, Yann LeCun et al. "Comparison of classifier methods: a case study in handwritten digit recognition." In Proceedings of the 12th IAPR International Conference on Pattern Recognition, Vol. 3-Conference C: Signal Processing (Cat. No. 94CH3440-5), vol. 2, pp. 77-82. IEEE, 1994.

Boutsidis, Christos, and Efstratios Gallopoulos. "SVD based initialization: A head start for nonnegative matrix factorization." Pattern recognition 41, no. 4 (2008): 1350-1362.

Brandt, Jakob, and Emil Lanzén. "A comparative review of SMOTE and ADASYN in imbalanced data classification." (2021).

Brownlee, Jason. *Imbalanced classification with Python: better metrics, balance skewed classes, cost-sensitive learning*. 2020.

Brunet, Jean-Philippe, Pablo Tamayo, Todd R. Golub, and Jill P. Mesirov. "Metagenes and molecular pattern discovery using matrix factorization." Proceedings of the national academy of sciences 101, no. 12 (2004): 4164-4169.

Buckland, Michael, and Fredric Gey. "The relationship between recall and precision." *Journal of the American society for information science* 45, no. 1 (1994): 12-19.

Burman, Prabir. "A comparative study of ordinary cross-validation, v-fold cross-validation and the repeated learning-testing methods." Biometrika 76, no. 3 (1989): 503-514.

Caldeira, Gregory A., and John R. Wright. "Amici curiae before the supreme court: who participates, when, and how much?." The Journal of Politics 52.3 (1990): 782-806.

Caldeira, Gregory A., and John R. Wright. "Organized interests and agenda setting in the US Supreme Court." American Political Science Review 82.4 (1988): 1109-1127.

Cavnar, William B., and John M. Trenkle. "N-gram-based text categorization." Proceedings of SDAIR-94, 3rd annual symposium on document analysis and information retrieval. Vol. 161175. 1994.

Chang, Yin-Wen, and Chih-Jen Lin. "Feature ranking using linear SVM." In Causation and prediction challenge, pp. 53-64. PMLR, 2008.

Chawla, Nitesh V., Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. "SMOTE: synthetic minority over-sampling technique." Journal of artificial intelligence research 16 (2002): 321-357.

Chen, Tianqi, and Carlos Guestrin. "Xgboost: A scalable tree boosting system." In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pp. 785-794. 2016.

Cicero, Marcus Tullius, *On Divination* (44 BC)

Cortes, Corinna, and Vladimir Vapnik. "Support-vector networks." Machine learning 20 (1995): 273-297.

Danescu-Niculescu-Mizil, Cristian, Lillian Lee, Bo Pang, and Jon Kleinberg. "Echoes of power: Language effects and power differences in social interaction." In Proceedings of the 21st international conference on World Wide Web, pp. 699-708. 2012.

Devins, Neal, and Lawrence Baum. "Split definitive: How party polarization turned the Supreme Court into a partisan court." The Supreme Court Review 2016, no. 1 (2017): 301-365.

Feldman, Adam, and Alexander Kappner. "Finding certainty in cert: An empirical analysis of the factors involved in Supreme Court certiorari decisions from 2001-2015." Vill. L. Rev. 61 (2016): 795.

Fernández, Alberto, Salvador Garcia, Francisco Herrera, and Nitesh V. Chawla. "SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary." Journal of artificial intelligence research 61 (2018): 863-905.

Fernández, Alberto, Salvador García, Mikel Galar, Ronaldo C. Prati, Bartosz Krawczyk, and Francisco Herrera. Learning from imbalanced data sets. Vol. 10. Cham: Springer, 2018.

Guyon, Isabelle, Jason Weston, Stephen Barnhill, and Vladimir Vapnik. "Gene selection for cancer classification using support vector machines." Machine learning 46 (2002): 389-422.

Hanretty, Chris. A court of specialists: Judicial behavior on the UK Supreme Court. Oxford University Press, USA, 2020.

He, Haibo, and Edwardo A. Garcia. "Learning from imbalanced data." IEEE Transactions on knowledge and data engineering 21, no. 9 (2009): 1263-1284.

He, Haibo, Yang Bai, Edwardo A. Garcia, and Shutao Li. "ADASYN: Adaptive synthetic sampling approach for imbalanced learning." In 2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence), pp. 1322-1328. Ieee, 2008.

He, Haibo, and Yunqian Ma, eds. "Imbalanced learning: foundations, algorithms, and applications." (2013).

Joachims, Thorsten. "Text categorization with support vector machines: Learning with many relevant features." European conference on machine learning. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998.

Karttunen, Lauri, Jean-Pierre Chanod, Gregory Grefenstette, and Anne Schille. "Regular expressions for language engineering." Natural Language Engineering 2, no. 4 (1996): 305-328.

Kastellec, Jonathan P., Jeffrey R. Lax, Michael Malecki, and Justin H. Phillips. "Polarizing the electoral connection: partisan representation in Supreme Court confirmation politics." The journal of politics 77, no. 3 (2015): 787-804.

Khalilia, Mohammed, Sounak Chakraborty, and Mihail Popescu. "Predicting disease risks from highly imbalanced data using random forest." BMC medical informatics and decision making 11 (2011): 1-13.

Lain, Corinna Barrett. "Deciding death." Duke LJ 57 (2007): 1. p 75

Lee, Daniel D., and H. Sebastian Seung. "Learning the parts of objects by non-negative matrix factorization." Nature 401, no. 6755 (1999): 788-791.

Leiman, Joan Maisel. "The Rule of Four." Colum. L. Rev. 57 (1957): 975.

Lilleberg, Joseph, Yun Zhu, and Yanqing Zhang. "Support vector machines and word2vec for text classification with semantic features." In 2015 IEEE 14th International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC), pp. 136-140. IEEE, 2015.

Lin, Yuxiao, Yuxian Meng, Xiaofei Sun, Qinghong Han, Kun Kuang, Jiwei Li, and Fei Wu. "Bertgcn: Transductive text classification by combining gcn and bert." arXiv preprint arXiv:2105.05727 (2021).

Linzer, Peter. "The meaning of certiorari denials." Colum L. Rev. 79 (1979): 1227.

López, Victoria, Alberto Fernández, and Francisco Herrera. "On the importance of the validation technique for classification with imbalanced datasets: Addressing covariate shift when data is skewed." Information Sciences 257 (2014): 1-13.

López, Victoria, Alberto Fernández, Salvador García, Vasile Palade, and Francisco Herrera. "An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics." Information sciences 250 (2013): 113-141.

Luhn, Hans Peter. "A statistical approach to mechanized encoding and searching of literary information." IBM Journal of research and development 1, no. 4 (1957): 309-317.

Martin, Andrew D., and Kevin M. Quinn. "Dynamic ideal point estimation via Markov chain Monte Carlo for the US Supreme Court, 1953–1999." Political analysis 10, no. 2 (2002): 134-153.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. "Distributed representations of words and phrases and their compositionality." Advances in neural information processing systems 26 (2013).

Miller, George A. "WordNet: a lexical database for English." Communications of the ACM 38, no. 11 (1995): 39-41.

Monroe, Burt L., Michael P. Colaresi, and Kevin M. Quinn. "Fightin' words: Lexical feature selection and evaluation for identifying the content of political conflict." Political Analysis 16, no. 4 (2008): 372-403;

O'callaghan, Derek, Derek Greene, Joe Carthy, and Pádraig Cunningham. "An analysis of the coherence of descriptors in topic modeling." Expert Systems with Applications 42, no. 13 (2015): 5645-5657.

Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. "Thumbs up? Sentiment classification using machine learning techniques." arXiv preprint cs/0205070 (2002).

Pascal, Blaise. "Pensées", edited and translated by Roger Ariew (Hackett, 2004)

Peirce, Charles Santiago Sanders. "Prolegomena to an apology for pragmaticism." The Monist (1906): 492-546.

Perry, Hersel W. Deciding to decide: agenda setting in the United States Supreme Court. Harvard University Press, 2009.

Picard, Richard R., and R. Dennis Cook. "Cross-Validation of Regression Models." Journal of the American Statistical Association 79, no. 387 (1984): 575–83. https://doi.org/10.2307/2288403.

Quinn, Kevin M., Burt L. Monroe, Michael Colaresi, Michael H. Crespin, and Dragomir R. Radev. "How to analyze political attention with minimal assumptions and costs." American Journal of Political Science 54, no. 1 (2010): 209-228.

Quinonero-Candela, Joaquin, Masashi Sugiyama, Anton Schwaighofer, and Neil D. Lawrence, eds. Dataset shift in machine learning. Mit Press, 2008.

Ramezan, A, Christopher, Timothy A. Warner, and Aaron E. Maxwell. "Evaluation of sampling and cross-validation tuning strategies for regional-scale machine learning classification." Remote Sensing 11, no. 2 (2019): 185.

Revesz, Richard L., and Pamela S. Karlan. "Nonmajority Rules and the Supreme Court." U. Pa. L. Rev. 136 (1987)

Robbins, Ira P. "Justice by the Numbers: The Supreme Court and the Rule of Four-Or Is It Five." Suffolk UL Rev. 36 (2002)

Saito, Takaya, and Marc Rehmsmeier. "The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets." *PloS one* 10, no. 3 (2015): e0118432.

Salton, Gerard, and Christopher Buckley. "Term-weighting approaches in automatic text retrieval." Information processing & management 24, no. 5 (1988): 513-523.

Sebastiani, Fabrizio. "Machine learning in automated text categorization." ACM computing surveys (CSUR) 34.1 (2002): 1-47.

Shan, Guogen. "Monte Carlo cross-validation for a study with binary outcome and limited sample size." BMC Medical Informatics and Decision Making 22, no. 1 (2022): 1-15.

Shao, Jun. "Linear model selection by cross-validation." Journal of the American statistical Association 88, no. 422 (1993): 486-494.

Sim, Yanchuan, Bryan Routledge, and Noah Smith. "The utility of text: the case of Amicus briefs and the Supreme Court." In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 29, no. 1. 2015.

Stephenson, Ryan. "Federal Circuit Cases Selection at the Supreme Court: An Empirical Analysis." Geo. LJ 102 (2013): 271.

Stern, Robert L., and Eugene Gressman. Supreme Court practice: jurisdiction, procedure, arguing and briefing techniques, forms, statutes, rules for practice in the Supreme Court of the United States. Bureau of national affairs, 1985.

Sulea, Octavia-Maria, Marcos Zampieri, Mihaela Vela, and Josef Van Genabith. "Predicting the law area and decisions of French Supreme Court cases." arXiv preprint arXiv:1708.01681 (2017).

Swets, John A. "Measuring the accuracy of diagnostic systems." *Science* 240, no. 4857 (1988): 1285-1293.

Tanenhaus, Joseph, et al. "The Supreme Court's certiorari jurisdiction: Cue theory." Judicial decision making 111 (1963).

Teger, Stuart H., and Douglas Kosinski. "The cue theory of Supreme Court certiorari jurisdiction: A reconsideration." The Journal of Politics 42.3 (1980): 834-846.

Ulmer, S. Sidney, William Hintze, and Louise Kirklosky. "The decision to grant or deny certiorari: Further consideration of cue theory." Law & Society Review 6.4 (1972): 637-643.

Ulmer, S. Sidney. "The decision to grant certiorari as an indicator to decision" on the merits"." Polity 4.4 (1972)

van de Bijl, Etienne, Jan Klein, Joris Pries, Sandjai Bhulai, Mark Hoogendoorn, and Rob van der Mei. "The Dutch Draw: constructing a universal baseline for binary prediction models." *arXiv preprint arXiv:2203.13084* (2022).

Vapnik, Vladimir. The nature of statistical learning theory. Springer science & business media, 1999.

Watson, Alan trans ed. The Digest of Justinian. University of Pennsylvania Press, 2011.

Xu, Qing-Song, and Yi-Zeng Liang. "Monte Carlo cross validation." Chemometrics and Intelligent Laboratory Systems 56, no. 1 (2001): 1-11.

2 On the Conceptualization of Environmental, Social, Governance (ESG) Issues in Distinct Corporate Discourse Channels: a Natural Language Processing Approach

Abstract

A major debate in contemporary corporate law and governance concerns the use of "Environmental, Social and Governance" (ESG) measures for the evaluation of corporate performance. An ESG-based view on corporate performance evaluation is contrasted with the "shareholder primacy" view of corporate purpose which stresses the importance of maximizing shareholder value, and thus, focuses on traditional measures of corporate performance such as stock price. Given the growing acceptance of ESG by corporations, a key question that emerges in this space is whether corporate actors are honest in their commitments to ESG. This paper proposes a natural language processing approach to examine this question. We hypothesize that honesty on the issue of ESG can be investigated by examining the consistency in the use of ESG related terms in distinct situational, communicative, and discursive contexts. We propose a comparison of corporate actors' language use in two separate corporate discourse channels, or what we call the "candid" and the "official" discourse channels. The "candid" corporate discourse channel contains earning conference call transcripts which record live communications between corporate actors, analysts and shareholders. The "official" corporate discourse channel contains official written corporate communication, namely 10-K and 10-Q disclosures and ESG reports. To this end we create a novel dataset consisting of the text from mandatory 10-K and 10-Q disclosures, earning call transcripts and ESG reports from corporations whose CEOs signed Business Roundtable's "Statement on the Purpose of a Corporation" in August 2019. We subsequently implement a synchronic lexical semantic change detection approach to quantify the variation in meaning of ESG related terms across the two discourse channels. The initial results indicate that ESG discourse is similar across the distinct discourse channels which suggests either a consistent engagement in ESG-related discourse by corporate actors or, alternatively, confirms the view that language management pays a fundamental role in corporate discourse, thus establishing that the underlying corporate discourse channels necessarily generate similar information.

Nemo potest personam diu ferre fictam;
ficta cito in naturam suam recidunt.

[No one can wear a mask for long;
Pretences soon fall back into their true nature.]

Seneca, *De Clementia* (55 AD)[134]

[134] Seneca, Lucius Annaeus, *On Mercy*, (55 AD)
translation taken from *Seneca: Moral and political essays*. Cambridge University Press, 1995.

2.1 Introduction

"Environmental, Social, Governance" (ESG) considerations and practices have arguably taken over the corporate world by a storm.[135] The term ESG refers to criteria for measuring corporate performance based on principles of ethical and socially responsible investment.[136] Under an ESG-based evaluation of corporate performance, environmental, social and governance factors become "the key points to be considered in the process of investment analysis and decision making."[137] More concretely, an ESG rating-based approach to corporate performance evaluation would focus on a company's sustainability performance, environmental impact and record, its treatment of employees, its governance practices, among other things.[138] This approach to corporate investment analysis and evaluation can be contrasted with other perspectives on evaluation of corporate performance, for instance, the traditional corporate finance metrics such as "stock price" or "price-to-earnings ratios,"[139] etc. Thus, according to Widywati's survey of the subject, ESG has been applied to "filter out non-ethical companies", but also, more recently, to "balance between punishing non-performing companies and rewarding best-performing companies"[140] from a corporate social responsibility perspective.

In the first part of this essay, I will be motivating the study by discussing the conceptualization, history and critiques of ESG. A key question in this space is whether corporate actors are honest in their commitments to ESG. In the second part, I will be discussing previous work in this space. In the third part I will be discussing the data and the separation of corporate communication into two distinct discourse channels, namely, the "candid" and the "official" discourse channels. In the fourth part it is argued that corporate honesty can be measured by examining the lexical semantic change and consistency of ESG and related terms across the two proposed discourse channels. We

---

[135] Kell, Georg. "The remarkable rise of ESG." *Forbes. com* 11 (2018).

[136] Li, Ting-Ting, Kai Wang, Toshiyuki Sueyoshi, and Derek D. Wang. "ESG: Research progress and future prospects." *Sustainability* 13, no. 21 (2021): 11663.
See also: Widyawati, Luluk. "A systematic literature review of socially responsible investment and environmental social governance metrics." *Business Strategy and the Environment* 29, no. 2 (2020): 619-637.

[137] Li, Ting-Ting, Kai Wang, Toshiyuki Sueyoshi, and Derek D. Wang. "ESG: Research progress and future prospects." *Sustainability* 13, no. 21 (2021): 11663.

[138] Landi, Giovanni, and Mauro Sciarelli. "Towards a more ethical market: the impact of ESG rating on corporate financial performance." *Social responsibility journal* 15, no. 1 (2018): 11-27.

[139] See generally: Damodaran, Aswath. *Corporate finance*. Hoboken: Wiley, 1996.

[140] Widyawati, Luluk. "A systematic literature review of socially responsible investment and environmental social governance metrics." *Business Strategy and the Environment* 29, no. 2 (2020): 619-637.
See also: Whelan, Tenise, U. Atz, T. Van Holt, and C. Clark. "Uncovering the Relationship by Aggregating Evidence from 1,000 Plus Studies Published between 2015–2020." *Online verfügbar unter https://www. stern. nyu. edu/sites/default/files/assets/documents/NYU-RAM_ESG-Paper_2021% 20Rev_0. pdf, abgerufen* 15 (2021): 2022.

propose applying word-embedding based synchronic lexical semantic change detection techniques to measure change in meaning of ESG across the candid and official discourse channels. Subsequently, we discuss the results and the limitations, thereafter proceeding to the conclusion.

2.2 Conceptualization, History and Growing Critiques of ESG

The ESG-based approach to corporate performance evaluation traces its origins to the so-called "stakeholder capitalism" movement which has its roots in the Davos Manifesto of 1973.[141] Stakeholder capitalism's proponents have argued that a company should prioritize the interests of its "stakeholders," namely, the company's employees, customers, suppliers, the environment, and the broader community at large, as opposed to solely focusing on the interests of the "shareholders" via the maximization of shareholder value. For instance, the Davos Manifesto 1973 states that "the purpose of professional management is to serve clients, shareholders, workers and employees, as well as societies, and to harmonize the different interests of the stakeholders."[142] In other words, under "stakeholderism," corporate success should not be defined solely in terms of share price, as is traditionally the case, but should also be based on satisfying the concerns of the company's various stakeholders. Accordingly, under the stakeholderist view, balancing the expectations and the needs of stakeholders becomes the main task of corporate decision-making, and thus, ESG becomes a means of measuring a company's performance along the environmental, social and governance dimensions.[143] The genesis of the term ESG can be traced back to a 2004 United Nations Global Compact working group report to the UN Secretary General Kofi Annan facilitated by Ivo Knoepfel titled "Who cares wins."[144] The report states that "sound corporate governance and risk management systems are crucial pre-requisites to successfully implementing policies and measures to address environmental and social challenges" and the term ESG "is a way of highlighting the fact that these three areas are closely interlinked."[145] In other words, ESG

[141] Davos Manifesto 1973: A Code of Ethics for Business Leaders, available at: https://www.weforum.org/agenda/2019/12/davos-manifesto-1973-a-code-of-ethics-for-business-leaders/
See also: Rock, Edward B. "For Whom is the Corporation Managed in 2020?: The Debate over Corporate Purpose." European Corporate Governance Institute-Law Working Paper 515 (2020): 20-16.
See also: Schwab, Klaus. "World economic forum." *Global Competitiveness Report (2014-2015)* (2015).
[142] Davos Manifesto 1973: A Code of Ethics for Business Leaders, available at: https://www.weforum.org/agenda/2019/12/davos-manifesto-1973-a-code-of-ethics-for-business-leaders/
[143] Schwab, Klaus. *Stakeholder capitalism: A global economy that works for progress, people and planet*. John Wiley & Sons, 2021.
[144] Knoepfel, Ivo. "Who Cares Wins: Connecting Financial Markets to a Changing World, UN Environment Programme." (2004).
[145] *Ibid.*

is a manifestation of stakeholderist views on corporate purpose with an emphasis on quantification of corporate impact via the use of ESG metrics and ratings.[146]

The alternative to "stakeholderism" and ESG is the so-called "shareholder primacy" view on corporate purpose, which conversely, argues for the prioritization of the interests of shareholders over the interests of stakeholders.[147] The most well-known representative of the shareholder primacy view was Milton Friedman who famously stated that "the social responsibility of business is to increase its profits,"[148] and thus, everything else should be subordinate to this fundamental purpose of a corporation to make money for its shareholders. Friedman's argument is that corporate actors' pursuit of "social responsibility" is essentially another way of spending "other people's money," and that "the stockholders or the customers or the employees could separately spend their own money on the particular action if they wished to do so."[149] Critics of the so-called "Friedman doctrine", which is another name for the "shareholder primacy" view, argue that such a singular focus on maximizing shareholder value can lead to myopic corporate decision-making that is solely focused on corporate profits and financial performance at the expense of other longer term considerations, such as social and environmental impacts of a corporation.[150] Accordingly, it can be said that the "shareholder primacy" and "stakeholderism" dichotomy lies at the core of the debate surrounding ESG.

Despite the well-intentioned nature of stakeholderism as an alternative model to corporate purpose, there is nevertheless a growing skepticism of stakeholderism and ESG-based corporate performance evaluation. Elon Musk, being perhaps the most famous critic of ESG from the business perspective, frequently posts his opinions on the subject on Twitter. For instance, in a May 18, 2022, tweet Musk stated that: "Exxon is rated top ten best in the world for environment, social & governance (ESG) by S&P 500, while Tesla didn't make the list! ESG is a scam. It has been weaponized by phony social justice warriors."[151] In a subsequent tweet, Musk continued, stating that S&P 500, which is a company that rates other companies based on their ESG scores and performance, has

---

[146] Dorfleitner, Gregor, Gerhard Halbritter, and Mai Nguyen. "Measuring the level and risk of corporate responsibility–An empirical comparison of different ESG rating approaches." Journal of Asset Management 16 (2015): 450-466.

[147] Friedman, Milton. "A Friedman Doctrine – the Social Responsibility of Business Is to Increase Its Profits." The New York Times, September 13, 1970. Available at: https://www.nytimes.com/1970/09/13/archives/a-friedman-doctrine-the-social-responsibility-of-business-is-to.html

[148] *Ibid.*

[149] *Ibid.*

[150] Mulligan, Thomas. "A critique of Milton Friedman's essay 'the social responsibility of business is to increase its profits'." *Journal of Business Ethics* 5 (1986): 265-269.
See also: Sprinkle, Geoffrey B., and Laureen A. Maines. "The benefits and costs of corporate social responsibility." *Business Horizons* 53, no. 5 (2010): 445.

[151] Elon Musk, 18 May 2023 tweet, 9:09 AM, available at: https://twitter.com/elonmusk/status/1526958110023245829?s=20

"lost their integrity."[152] The simple point of Musk's tweets was to demonstrate that ESG metrics cannot accurately represent the "greenness" of a business, otherwise, an oil company like Exxon, which is a known polluter with a formidable pollution track record,[153] should not be ranked higher on ESG than Tesla, a company that manufactures electric vehicles.

Recent anti-ESG legislative efforts in Texas seem to suggest a conservative bent to ESG skepticism.[154] However, it is worth pointing out that there is also a rich history of leftist critiques of corporate social responsibility. For instance, Weinstein in his study on trade unionism in the United States in the early 20th century argues that the acceptance of "socially responsible" business practices was merely a means of combatting the rise of socialism in the United States,[155] stating that, "throughout the Progressive Era a primary concern of corporation and financial leaders associated with the National Civic Federation, as well as of almost all major political figures, was the promotion of social responsibility," which effectively "strengthen[ed] conservative trade unionism as epitomized by Samuel Gompers against the challenges of Socialists within the American Federation of Labor."[156] Furthermore, Weinstein quotes Louis D. Brandeis, who stated that "the trade unions also stand as a strong bulwark against the great wave of socialism."[157] A more recent leftist critique comes from Žižek, who gives an example of Starbucks Coffee being comparatively more expensive than other coffees due to what he calls a "cultural surplus."[158] According to Žižek, what the Starbucks customer is "really buying is the "coffee ethic" which includes care for the environment, social responsibility towards the producers."[159] From this perspective, corporate ethics can be interpreted as nothing more than a marketing strategy to make the consumer feel less guilty about being a consumerist. Thus, according to Žižek, the knowledge that consumption is "ethical" is itself the product that is being sold, and the customer is willing to pay a bit more for a Starbucks Coffee because it also provides a guilt-free conscience.[160] Indeed, the view that

---

[152] Elon Musk, 18 May 2023 tweet, 9:18 AM, available at: https://twitter.com/elonmusk/status/1526960512231153664?s=20

[153] Boehm, Paul D., Gregory S. Douglas, William A. Burns, Paul J. Mankiewicz, David S. Page, and A. Edward Bence. "Application of petroleum hydrocarbon chemical fingerprinting and allocation techniques after the Exxon Valdez oil spill." *Marine Pollution Bulletin* 34, no. 8 (1997): 599-613.
See also: Xia, Yuqiang, and Michel C. Boufadel. "Lessons from the Exxon Valdez oil spill disaster in Alaska." *Disaster Advances* 3, no. 4 (2010): 270-273.

[154] "Governor Abbott Denounces ESG Standards Harmful to U.S. Energy Sector." Office of the Governor of Texas, Greg Abbott, available at: https://gov.texas.gov/news/post/governor-abbott-denounces-esg-standards-harmful-to-u.s-energy-sector

[155] Weinstein, James. "The corporate ideal in the liberal state: 1900-1918." (1969).

[156] *Ibid.* p 117

[157] *Ibid.* p 17

[158] Žižek, Slavoj. *First as tragedy, then as farce*. Verso, 2009.

[159] *Ibid.*

[160] *Ibid.*

ESG is either illusory or just another marketing ploy is becoming more prevalent as terms like "green-washing", which refers to the act of "misleading consumers about [a corporation's] environmental performance or the environmental benefits of a product or service,"[161] have arguably entered the vernacular. A recent NY Times article discusses the rapid rise of ESG on Wall Street of all places, stating that "ESG fund makeovers have become the trend du jour on Wall Street: BlackRock, J.P. Morgan, Morgan Stanley, HSBC, WisdomTree, Putnam and MassMutual have all done it."[162] Some academics have also questioned ESG from various disciplinary angles, such as law, finance, economics, etc.[163] Additionally, ESG and related issues are currently being targeted by regulators. For instance, on March 21st, 2022, the Securities Exchange Commission (SEC) has proposed "Rules to Enhance and Standardize Climate-Related Disclosures for Investors."[164] Two months later, on May 25th, 2022, the SEC proposed another set of rules to "Enhance Disclosures by Certain Investment Advisers and Investment Companies About ESG Investment Practices."[165] These proposals by the SEC clearly indicate a willingness to regulate the ESG space. Overall, a simple question thus emerges throughout this discussion: are corporations honest in their commitments to ESG?

[161] Delmas, Magali A., and Vanessa Cuerel Burbano. "The drivers of greenwashing." *California management review* 54, no. 1 (2011): 64-87.
[162] Goldstein, Matthew, and Emily Flitter. "Cracking down on a Wall Street Trend: E.S.G. Makeovers." The New York Times, September 17, 2022.
https://www.nytimes.com/2022/09/17/business/dealbook/esg-wall-street.html.
[163] Raghunandan, Aneesh, and Shivaram Rajgopal. "Do socially responsible firms walk the talk?." *Available at SSRN 3609056* (2022).
See also: Berk, Jonathan, and Jules H. Van Binsbergen. "The impact of impact investing." (2021).
See also: Strine Jr, Leo E., Justin L Brooke, Kyle M. Diamond, and Derrick L. Parker Jr. "It's Time to Focus on the "G" in ESG." *Harvard Business Review, November* 18 (2022).
See also: Yu, Ellen Pei‐yi, Christine Qian Guo, and Bac Van Luu. "Environmental, social and governance transparency and firm value." *Business Strategy and the Environment* 27, no. 7 (2018): 987-1004.
[164] Rules to Enhance and Standardize Climate-Related Disclosures for Investors. "Press Release." SEC, March 21, 2022.
Available at: https://www.sec.gov/news/press-release/2022-46
[165] Enhance Disclosures by Certain Investment Advisers and Investment Companies About ESG Investment Practices "Press Release." SEC, May 25, 2022.
Available at: https://www.sec.gov/news/press-release/2022-92

2.3 Related Work

The present investigation builds upon several strands of research. First, we investigate work in law, business ethics, ESG and stakeholderism, corporate governance, sustainability, and related issues. Second, we investigate natural language processing approaches with a focus on adopting methods for synchronic lexical semantic change detection across domains as a means of measuring corporate honesty on ESG.

When it comes to text analytic research on corporate discourse Li's work offers a foundational survey on corporate textual data and methodologies.[166] Notably, Li proposes that one of the areas of further research should indeed be the study of "different [corporate] discourse channels", noting that "most existing papers focus on one specific textual discourse channel."[167] As will be discussed later, we adopt Li's suggestion by proposing a comparison of two corporate discourse channels, namely the "candid" and "official" discourse channels. Li also points out some important limitations when it comes to corporate textual data, namely that even in "different discourse channels, there are likely influences of similar economic factors."[168] In other words, a key problem of studying corporate discourse is that sophisticated corporate actors likely generate similar linguistic data regardless of which channel they utilize for communication. This view echoes Bergelia and Harris' systematic linguistic study of corporate discourse, where the importance of "language management" in corporate organizational settings is emphasized.[169] For instance, Bergelia and Harris note that "insist[ing] on sense-making is to acknowledge its fundamental role in the interpretation of organizational dis-order where meetings perform the function of public fora that enable individuals to confront and adjust to multiple realities, i.e. weave the threads of interpersonal and social coherence."[170] We interpret this to mean that corporate realities necessarily demand a unified, coherent and non-contradictory organizational perspective which will mean that language use will likely be similar regardless of the discourse channel utilized for communication, and this is indeed what our results seem to suggest. In short, the study of language in different corporate discourse channels is an important area of research, but one of the key issues is that the language generated in these channels will likely be similar, because communication in various corporate discourse channels is essentially, in the words of Li, influenced by "similar economic factors."[171]

---

[166] Li, Feng. "Textual analysis of corporate disclosures: A survey of the literature." *Journal of accounting literature* 29 (2010): 143.
[167] *Ibid.*
[168] *Ibid.*
[169] Bargiela, Francesca, and Sandra J. Harris. *Managing language: The discourse of corporate meetings*. Vol. 44. John Benjamins Publishing, 1997.
[170] *Ibid.* p 57
[171] Li, Feng. "Textual analysis of corporate disclosures: A survey of the literature." *Journal of accounting literature* 29 (2010): 143.

When it comes to corporate honesty on ESG, Raghanundan and Rajgopal investigate whether the "socially responsible walk the walk."[172] In this study, the "socially responsible" were defined as those corporations which signed the August 2019 Business Roundtable declaration "On the Purpose of the Corporation." Using a propensity score matching technique to match non-Business Roundtable firms with Business Roundtable signatory firms, Raghanundan and Rajgobal show that Business Roundtable signatory firms "report higher rates of environmental and labor-related compliance violations" when compared to their matched peer firms in the non-Business Roundtable group, suggesting that the socially responsible despite talking the talk, do not necessarily walk the walk when it comes to their ESG commitments, thus providing the evidence of corporate dishonesty on the issue of ESG.[173] In our study, we adopt Raghanundan and Rajgopal's definition of "socially responsible corporations" as those corporations whose CEOs signed the August 2019 Business Roundtable statement "On the Purpose of a Corporation." Thus, for our data collection efforts, we focus solely on these Business Roundtable signatory firms and subsequently, we compare the text in their earning call transcripts, which we call the "candid" corporate discourse channel, with the text in their 10-K, 10-Q and ESG report documents, or the "official" corporate discourse channel.

Another recent study by Christensen, Serafim and Sekochi studies ESG disclosures, data and metrics used by ESG rating agencies, finding that more ESG disclosures released by corporations lead to less consistent ratings by rating agencies, which suggests, somewhat counterintuitively, that when it comes to ESG, more information leads to less consistency in ESG ratings.[174] The authors argue that this uncertainty in ratings shows that ESG is experiencing an "early stage of institutional innovation," signifying an uncertain nature of the ESG phenomenon.[175]

With regards to the related concept of sustainability, relevant research comes from Laine's investigation of the meaning of the term "sustainable development" in Finnish corporate discourse.[176] Laine employs critical discourse methodology which involves closely reading corporate disclosure documents in question and identifying the emerging discursive practices and patterns.[177] After studying 80 annual reports and 25 other reports, Laine identifies a rhetoric of "weak sustainability" in business actors' discourse practices which "presents [sustainable development] as a sort of a holy grail, which will simultaneously endow society with further economic growth, environmental protection

[172] Raghunandan, Aneesh, and Shivaram Rajgopal. "Do socially responsible firms walk the talk?." *Available at SSRN 3609056* (2022).
[173] *Ibid.*
[174] Christensen, Dane M., George Serafeim, and Anywhere Sikochi. "Why is corporate virtue in the eye of the beholder? The case of ESG ratings." *The Accounting Review* 97, no. 1 (2022): 147-175.
[175] *Ibid.*
[176] Laine, Matias. "Meanings of the term 'sustainable development' in Finnish corporate disclosures." In *Accounting Forum*, vol. 29, no. 4, pp. 395-413.
[177] *Ibid.*

and social improvements, with little or no trade-offs."[178] The "weak sustainability view" is contrasted with the "strong sustainability" view, which is distinguished by its emphasis on urgency, difficult trade-offs and the necessity of immediate action on part of corporations to meet their sustainability objectives.[179] The Laine study demonstrates the difficulty and the time-consuming nature of close reading techniques which invites us to explore novel natural language processing methods for the investigation of the question of corporate honesty on ESG.

From a business theory and ethics perspective, Camilleri[180] offers a survey of the theories behind integrated reporting, of which voluntary corporate ESG disclosures and reports can be seen as an example. Camilleri examines the various theories behind why corporations would voluntarily communicate their commitments in their integrated reports: the agency theory, stewardship theory, institutional theory, legitimacy theory, and the isomorphism/isopraxism theory.[181] The legitimacy theory seems to be the most relevant to our present work, which states that, "organizations prepare integrated reports in an attempt to maintain or repair their legitimacy among stakeholders."[182] Thus, according to Camilleri, "responsible organizations become legitimate entities, particularly if they comply with relevant societal rules and norms."[183] Camilleri cites Suchman who argues that legitimacy can also be conceptualized as an operational resource.[184] Conceptualizing corporate legitimacy as a resource will likely necessitate special attention to consistency on the issue of ESG throughout the various communication channels, especially given the fact that environmental issues resonate with the public at large.[185]

From a corporate law and governance perspective, Strine argues that corporate law in the US is fundamentally shareholder-centric, and that the only way to satisfy stakeholder interests is to adopt statutes "giving [stakeholders] enforceable rights that they can wield,"[186] otherwise, any discussions on stakeholderism, commitments to ESG, and corporate social responsibility are legally inconsequential. Lund and Pollman also argue that ESG considerations are likely impossible to satisfy given that US corporate law is

[178] *Ibid.*
[179] *Ibid.*
[180] Camilleri, Mark Anthony. "Theoretical insights on integrated reporting: The inclusion of non-financial capitals in corporate disclosures." *Corporate Communications: An International Journal* 23, no. 4 (2018): 567-581.
[181] *Ibid.*
[182] *Ibid.*
[183] *Ibid.*
[184] Suchman, Mark C. "Managing legitimacy: Strategic and institutional approaches." Academy of management review 20, no. 3 (1995): 571-610.
[185] Pérez, Lucy, Vivian Hunt, Hamid Samandari, Robin Nuttall, and Krysta Biniek. "Does ESG really matter—and why." *McKinsey Quarterly* (2022).
[186] Strine Jr, Leo E. "The dangers of denial: The need for a clear-eyed understanding of the power and accountability structure established by the Delaware general corporation law." *Wake Forest L. Rev.* 50 (2015): 761.

fundamentally based on ideas of shareholder primacy.[187] Crucially for present research, Lund and Pollman argue that "culture, the final component of the corporate governance machine, may be the most influential of all" when it comes to the ESG question,[188] which we take as an invitation to study the language of corporate actors on ESG issues.

When it comes to natural language processing based research on ESG and related issues, Jaworska and Nanda applied topic modeling assisted discourse analysis approaches to study the themes and patterns in a corpus made up of corporate discourse on social responsibility.[189] In another study, Jaworska utilizes traditional corpus linguistic techniques such as word frequencies and collocations to investigate the patterns in discussions on climate change.[190] Our methodological contribution is the application of word embedding based lexical semantic change detection techniques to the question of corporate honesty on ESG.[191]

Hamilton et al.'s work is the main methodological framework for our proposed NLP approach.[192] In their paper, Hamilton et al. study diachronic lexical semantic change by constructing word2vec word embedding models for different time periods and aligning these models using orthogonal Procrustes Alignment techniques.[193] Using cosine distance measurements for an individual word's vector representation between different time periods allows for a quantitative measurement of lexical semantic change over time.[194] Recently, Schlechtweg et al. have demonstrated that the aforementioned technique of building word embedding models for distinct time periods, aligning them and

---

[187] Lund, Dorothy S., and Elizabeth Pollman. "The corporate governance machine." *Colum. L. Rev.* 121 (2021): 2563.
[188] *Ibid.*
[189] Jaworska, Sylvia, and Anupam Nanda. "Doing well by talking good: A topic modelling-assisted discourse study of corporate social responsibility." *Applied Linguistics* 39, no. 3 (2018): 373-399.
[190] Jaworska, Sylvia. "Change but no climate change: Discourses of climate change in corporate social responsibility reporting in the oil industry." *International Journal of Business Communication* 55, no. 2 (2018): 194-219.
See generally: Jaworska, Sylvia. "Corporate discourse." (2020).
[191] Hamilton, William L., Jure Leskovec, and Dan Jurafsky. "Diachronic word embeddings reveal statistical laws of semantic change." *arXiv preprint arXiv:1605.09096* (2016).
Schlechtweg, Dominik, Anna Hätty, Marco Del Tredici, and Sabine Schulte im Walde. "A wind of change: Detecting and evaluating lexical semantic change across times and domains." *arXiv preprint arXiv:1906.02979* (2019).
[192] Hamilton, William L., Jure Leskovec, and Dan Jurafsky. "Diachronic word embeddings reveal statistical laws of semantic change." *arXiv preprint arXiv:1605.09096* (2016)..
[193] *Ibid.*
See also: Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. "Distributed representations of words and phrases and their compositionality." *Advances in neural information processing systems* 26 (2013).
Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. "Efficient estimation of word representations in vector space." *arXiv preprint arXiv:1301.3781* (2013).
[194] *Ibid.*

subsequently measuring the cosine distances for given tokens offers "impressively high performance and robustness."[195] They confirm this systematically by testing this method on datasets with known lexical semantic changes and divergences. They also provide incredibly useful code which we use for this research.[196]

Schlechtweg et al. also propose a second type of lexical semantic change detection research, namely the study of synchronic, rather than diachronic lexical semantic change, which they define as "NLP research with a focus on how the meanings of words vary across domains or communities of speakers."[197] Practically speaking, measuring synchronic semantic change would involve aligning word embedding models across different domains, rather than across different time periods. We believe that this approach seems perfectly suited for studying semantic differences regarding ESG in distinct corporate discourse channels.

According to Schlechtweg et al.[198] "the only notable work explicitly measuring across domain meaning shifts is Ferrari et al. paper,"[199] which develops a simple method for finding differences in word meanings across distinct domains, but without utilizing Procrustes Alignment techniques introduced in the Hamilton et al. paper.[200] For instance, using Ferrari's methodology, one can synchronically examine the emerging nearest neighbors for a word like "code" in the computer science domain, namely, words like "compile, executable, runtime, statically", and compare these words to the nearest neighbor words for "code" in the sports domain, namely, words like "rule, regulation, guideline."[201] Again, unlike Hamilton et al., Ferrari et al. do not implement orthogonal Procrustes alignment between domain-specific word embedding vector spaces, but rather use a "word injection method," which despite its intuitiveness was "consistently outperformed by orthogonal Procrustes" in the Schlechtweg et al. study.[202] This limitation is the main reason we prefer using Hamilton et al.'s orthogonal Procrustes based

---

[195] Schlechtweg, Dominik, Anna Hätty, Marco Del Tredici, and Sabine Schulte im Walde. "A wind of change: Detecting and evaluating lexical semantic change across times and domains." *arXiv preprint arXiv:1906.02979* (2019).
[196] *Ibid.*
[197] *Ibid.*
[198] *Ibid.*
[199] Ferrari, Alessio, Beatrice Donati, and Stefania Gnesi. "Detecting domain-specific ambiguities: an NLP approach based on Wikipedia crawling and word embeddings." In *2017 IEEE 25th International Requirements Engineering Conference Workshops (REW)*, pp. 393-399. IEEE, 2017.
[200] Hamilton, William L., Jure Leskovec, and Dan Jurafsky. "Diachronic word embeddings reveal statistical laws of semantic change." *arXiv preprint arXiv:1605.09096* (2016).
[201] Ferrari, Alessio, Beatrice Donati, and Stefania Gnesi. "Detecting domain-specific ambiguities: an NLP approach based on Wikipedia crawling and word embeddings." In *2017 IEEE 25th International Requirements Engineering Conference Workshops (REW)*, pp. 393-399. IEEE, 2017.
[202] *Ibid.*

approach to lexical semantic change detection, but nevertheless, for future work, we intend to also apply Ferrari's methodology to our research question.

To conclude, to the best of our knowledge, no other research has been done on examining ESG and related issues in corporate discourse using word embedding based synchronic lexical semantic change detection techniques.

2.4 Data

For our corpus we focus on public companies whose executives signed the Business Roundtable's "Statement on the Purpose of a Corporation" in August 2019.[203] Despite its brevity, this statement is a significant restatement of corporate purpose from a stakeholderist perspective. For instance, the press release explicitly states that the Business Roundtable Statement "moves away from shareholder primacy and includes commitment to all stakeholders."[204] Since August 2019, more company executives have been signing on to this statement, but we investigate only the initial signatories.

As mentioned earlier in the literature review, the Business Roundtable Statement signatory companies were used by Raghanundan and Rajgopal as a representative sample of "socially responsible firms."[205] However, unlike the Raghanunda and Rajgopal study, we do not use a propensity score matching technique to find matching peer companies. Rather, we want to compare these Business Roundtable Signatory companies' discourse as manifested in two distinct corporate discourse channels in order to explore their linguistic consistency and honesty on the topic of ESG. We also exclude private companies that are signatories from the sample as they do not have the same mandatory disclosure requirements as public companies, and thus do not have the same text data availability as public companies. This leaves us with 146 companies, and we collect corporate discourse data from 2017 to 2021 as these periods coincide with the rise of ESG.[206]

---

[203] Business Roundtable - Purpose of a Corporation. First published, August 19, 2019.
Available at: https://purpose.businessroundtable.org/

[204] "Business Roundtable Redefines the Purpose of a Corporation to Promote 'an Economy That Serves All Americans. Updated Statement Moves Away from Shareholder Primacy, Includes Commitment to All Stakeholders"
Available at:
https://www.businessroundtable.org/business-roundtable-redefines-the-purpose-of-a-corporation-to-promote-an-economy-that-serves-all-americans

[205] Raghunandan, Aneesh, and Shivaram Rajgopal. "Do socially responsible firms walk the talk?." *Available at SSRN 3609056* (2022).

[206] Kell, Georg. "The remarkable rise of ESG." *Forbes. com* 11 (2018).
We do not study discourse after 2022 because we were collecting data in 2022.

2.4.1 Candid and Official Corporate Discourse Channels

Our aim is to investigate the question of corporate honesty on the issue of ESG by examining synchronic lexical semantic change of ESG-related terms across the two corporate discourse channels for Business Roundtable Signatory corporations. We thus define the two corporate discourse channels as follows: (1) the "candid" corporate discourse channel, which contains the earning conference call transcripts, and (2) the "official" corporate discourse channel, which contains quarterly 10-Q and yearly 10-K mandatory disclosures supplemented with voluntary corporate ESG reports.

To get the data for our "candid" corporate discourse channel, we download earning conference call transcript .pdf files from FactSet,[207] and subsequently, we extract the text of these files using the tika pdf parser for Python.[208] During earning conference calls corporate executives update their shareholders on the corporation's actions in a given quarter and subsequently proceed to answer questions which are usually asked by business analysts. As one can imagine, earning calls discuss a wide range of corporate issues and topics, ESG being one of them. We assume that the earning call transcripts are more "candid" because during earning calls corporate representatives speak live over a phone in a group meeting setting, and indeed, as Li points out "statements made by managers in the Q&A section of the conference calls are more spontaneous and thus less likely to be influenced by staged preparation."[209] Chen, Hollander and Law also examined the influence of earning calls on stock price, finding that lack of questions during earning conference calls leads to stock price decreases, therefore, there is an incentive to answer questions.[210]

To get the data for our "official" corporate discourse channel we use the "SEC EDGAR" text crawler.[211] This crawler is convenient as it scrapes only the text of the

---

[207] Earning conference call transcript data were downloaded with the help of Research Assistants from: https://www.factset.com/
Here, I would like to extend my immense gratitude to the exceptional team of Research Assistants (RAs) from UC Berkeley's "Data Science Discovery Program" – namely, Angela Feng, Anjing Li, Genie Lee, Nianyao Du and Shunsuke Kinoshita. The Data Science Discovery is a unique data science-focused research incubator program at UC Berkeley's College of Computing, Data Science and Society which pairs talented and exceptional undergraduate students with scholars and researchers. For two semesters, the RAs played an instrumental role in helping with this project by assisting with the many difficult and time-consuming aspects of this research, namely developing the dataset by downloading the .pdf files from Factset and related tasks, crafting efficient code for the conversion of .pdf files to text, data cleaning, and providing preliminary analysis on the large amounts of textual data, among many other things.
[208] https://github.com/chrismattmann/tika-python
[209] Li, Feng. "Textual analysis of corporate disclosures: A survey of the literature." *Journal of accounting literature* 29 (2010): 143.
[210] Chen, Shuping, Stephan Hollander, and Kelvin Law. "In search of interaction." *Available at SSRN 2449341* (2016).
[211] https://github.com/alions7000/SEC-EDGAR-text

mandatory 10-K and 10-Q disclosures rather than the data tables with various accounting measures. Of particular significance in 10-Ks and 10-Qs are "Management's Discussion and Analysis of Financial Condition and Results of Operations" (MD&A) section,[212] which includes narrative statements about a company's operations in a given time period, and the "Risk Factors" section.[213] Both 10-Ks and 10-Qs contain this section, but the difference is that the information in 10-Qs is unaudited.[214] Mandatory disclosures tend to be lengthy and contain many redundancies and repetitions.[215] As a result, in our inspection of the collected mandatory disclosures we find that informative discussions on ESG are lacking, likely because ESG is still a novel and uncertain institutional concept.[216] To overcome a lack of useful information on ESG in 10-Ks and 10-Qs, we go for another official source of corporate information that companies have, namely their voluntary ESG reports. In these ESG reports, companies provide extensive detail on their ESG efforts, but unlike mandatory reports, they are published voluntarily and are thus not subject to the same legal scrutiny as mandatory reports. This is an important limitation when it comes to constructing the "official" discourse channel because the mandatory and the voluntary documents, despite both being official corporate communication, do not have the same incentives. For future work, as the concept of ESG becomes more settled in law, and thus, discussed in greater detail, we intend to focus solely on mandatory disclosures. A possible solution can also involve excluding 10-Ks and 10-Qs entirely, however, the value of 10-Qs and 10-Ks for the present research becomes apparent when we take into account non-ESG topics which will be used as a baseline for comparison of ESG-related lexical semantic change across corporate discourse channels. Thus, for instance, when discussing "tax" or "investment" topics we can be confident that the meaning of these words across the discourse channels will likely remain the same as these issues are uncontroversial. As with earning call transcript .pdfs, we similarly use the tika pdf parser

---

[212] Brown, Stephen V., and Jennifer Wu Tucker. "Large‐sample evidence on firms' year‐over‐year MD&A modifications." *Journal of Accounting Research* 49, no. 2 (2011): 309-346.

[213] Campbell, John L., Hsinchun Chen, Dan S. Dhaliwal, Hsin-min Lu, and Logan B. Steele. "The information content of mandatory risk factor disclosures in corporate filings." *Review of Accounting Studies* 19 (2014): 396-455.

[214] Chiu, Tiffany, Feiqi Huang, Yue Liu, and Miklos A. Vasarhelyi. "The impact of non-timely 10-Q filings and audit firm size on audit fees." *Managerial Auditing Journal* 33, no. 5 (2018): 503-516.

See also: Kim, Hyogon, Eunmi Lee, and Donghee Yoo. "Do SEC filings indicate any trends? Evidence from the sentiment distribution of forms 10-K and 10-Q with FinBERT." *Data Technologies and Applications* ahead-of-print (2023).

[215] Cazier, Richard A., and Ray J. Pfeiffer. "Why are 10-K filings so long?." *Accounting Horizons* 30, no. 1 (2016): 1-21.

See also: Lewis, Craig, and Steven Young. "Fad or future? Automated analysis of financial text and its implications for corporate reporting." *Accounting and Business Research* 49, no. 5 (2019): 587-615.

[216] Christensen, Dane M., George Serafeim, and Anywhere Sikochi. "Why is corporate virtue in the eye of the beholder? The case of ESG ratings." *The Accounting Review* 97, no. 1 (2022): 147-175.

for the corporate ESG reports.[217] Unfortunately, the parsing is far from ideal due to the numerous figures and pictures in these reports, but nevertheless, the parser is able to preserve much of the textual information.

Before proceeding further, it is important to acknowledge the fundamental limitations of the proposed separation of corporate communication into these discourse channels and the underlying corporate discourse data. Firstly, it is important to recognize that the "official" corporate discourse channel is not well-defined as we have to supplant mandatory disclosures with voluntary ESG reports due to a lack of informative ESG discussions in mandatory disclosures. Further work here is thus necessary. Secondly, corporate managers tend to be sophisticated actors who are incredibly careful about their language use.[218] In fact, as Bargiela and Harris point out "interactants [in business meetings] make sense of organizational life and build coherence patterns pragmatically through language use."[219] Accordingly, establishing coherence in an organizational setting across different communicative contexts is a fundamental part of a corporate managerial job, and it can indeed be said that, the language of business is essentially the "business of language."[220] Therefore, because linguistic coherence in corporate communication is so important to corporate actors, it is likely that the proposed separation into "official" and "candid" discourse channels is unfounded. Again, as Li pertinently points out, in "different discourse channels, there are likely influences of similar economic factors."[221]

A further limitation of the underlying data is that it is incredibly noisy. Perhaps unsurprisingly, most companies discuss only company-specific issues. For instance, Apple will discuss its new product both in the earning conference call transcripts and the disclosures, while American Airlines will discuss upgrading its fleet, etc. Finding common issues and topics between various companies has proven to be an incredibly difficult task, necessitating the use of topic modeling techniques which will be discussed later. Furthermore, there is an inherent difficulty in comparing spoken language as represented in the earning conference call transcripts with written language as represented in the official corporate documents, and thus, an added level of noise exists there.

---

[217] https://github.com/chrismattmann/tika-python

[218] Bargiela, Francesca, and Sandra J. Harris. *Managing language: The discourse of corporate meetings*. Vol. 44. John Benjamins Publishing, 1997.
See also: Amernic, Joel, and Russell Craig. CEO-speak: The language of corporate leadership. McGill-Queen's Press-MQUP, 2006.

[219] Bargiela, Francesca, and Sandra J. Harris. *Managing language: The discourse of corporate meetings*. Vol. 44. John Benjamins Publishing, 1997.

[220] *Ibid.*

[221] Li, Feng. "Textual analysis of corporate disclosures: A survey of the literature." *Journal of accounting literature* 29 (2010): 143.

2.5 Methodology

This paper seeks to examine corporate honesty on ESG by examining corporate communication and comparing language use in distinct corporate discourse channels. Getting at issues of honesty and deception is a difficult task, especially considering the previous discussions on corporate actors' "language management" and sophistication.[222] Our research endeavor is further complicated by the fact that we're dealing only with textual data. For instance, studies of deception and subjective attitude have been effectively explored with the help of acoustic data.[223] For future work, a study of verbal and non-verbal signals in corporate actors' voices can be more beneficial for exploring questions of honesty in corporate discourse, especially in light of the earning call audio data being recently made available.[224]

Following Kenneth R. Hammond's approach to the evaluation of judgements we can define dishonesty as the inability to meet the coherence theory of truth.[225] Hammond

---

[222] Feely, Alan J., and Anne‐Wil Harzing. "Language management in multinational companies." *Cross Cultural Management: an international journal* 10, no. 2 (2003): 37-52.
See also: Bargiela, Francesca, and Sandra J. Harris. *Managing language: The discourse of corporate meetings*. Vol. 44. John Benjamins Publishing, 1997.
See also: Amernic, Joel, and Russell Craig. CEO-speak: The language of corporate leadership. McGill-Queen's Press-MQUP, 2006.

[223] Graciarena, Martin, Elizabeth Shriberg, Andreas Stolcke, Frank Enos, Julia Hirschberg, and Sachin Kajarekar. "Combining prosodic lexical and cepstral systems for deceptive speech detection." In 2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings, vol. 1, pp. I-I. IEEE, 2006.
Voigt, Rob, Nicholas P. Camp, Vinodkumar Prabhakaran, William L. Hamilton, Rebecca C. Hetey, Camilla M. Griffiths, David Jurgens, Dan Jurafsky, and Jennifer L. Eberhardt. "Language from police body camera footage shows racial disparities in officer respect." *Proceedings of the National Academy of Sciences* 114, no. 25 (2017): 6521-6526.
See generally: Gregory Jr, Stanford W., and Stephen Webster. "A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perceptions." *Journal of personality and social psychology* 70, no. 6 (1996): 1231.
See also: Danescu-Niculescu-Mizil, Cristian, Lillian Lee, Bo Pang, and Jon Kleinberg. "Echoes of power: Language effects and power differences in social interaction." In *Proceedings of the 21st international conference on World Wide Web*, pp. 699-708. 2012.
However, see also: Galasinski, Dariusz. *The language of deception: A discourse analytical study*. Sage Publications, 2000.

[224] Li, Jiazheng, Linyi Yang, Barry Smyth, and Ruihai Dong. "Maec: A multimodal aligned earnings conference call dataset for financial risk prediction." In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pp. 3063-3070. 2020.
See also: Qin, Yu, and Yi Yang. "What you say and how you say it matters: Predicting stock volatility using verbal and vocal cues." In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 390-401. 2019.

[225] Hammond, Kenneth R. *Human judgment and social policy: Irreducible uncertainty, inevitable error, unavoidable injustice*. Oxford University Press, USA, 1996.
See also: Hammond, Kenneth R. "Coherence and correspondence theories in judgment and decision making." (2000).

defines coherence competence as the evaluation of "the consistency of the elements of the person's judgment."[226] According to Hammond, "coherence theories demand that the facts "hang together," that they tell a good story, one that is plausible and compelling because there are no discordant elements. Thus, coherence is often used as a criterion for whether we choose to believe a story or a theory."[227] Indeed, as Blair et al. point out, these methods have been "applied to the deception detection literature as strategies for distinguishing honest messages from deceptive messages."[228]

As an illustration, let us imagine the following scene: a student in a university tells the Professor the following statement after a lecture: (1) "I'm really enjoying your *class*! It's super interesting!" Judging this statement solely from a lexical perspective we can state that the motivations for making this statement could arise from several factors. Firstly, perhaps the student really likes the class, or perhaps they need to get a good grade, or they think that by being nice to the professor they will be perceived as a good student, or perhaps they seek to get a letter of recommendation for some application after the class is over, etc. But the intentions for uttering the statement above become clearer once we observe the student talking to their friends later that day, saying the following: (2) "Actually, the *class* is really boring. I literally fell asleep!" We can clearly observe that an obvious case of dishonesty is at play here. In the example above the two statements do not meet the coherence criterion in that the first statement of the student does not match up with the second statement uttered after the lecture, and the student's narration is thus demonstrably lexically discordant.

What's more interesting however is that the words surrounding the italicized word "*class*" in the statements above are inconsistent when uttered in distinct contextual domains and social situations. It is possible that this feature of the coherence-based definition of honesty can likely be used linguistically. In essence: when one switches "domains", for example, as when the student talked to the professor in statement (1) to when the student talked to their friends in statement (2), it is clear that different contextual words around the word "*class*" are generated depending on the situational contexts, manifested as a lexical inconsistency surrounding the term "*class*." In more concrete terms, in statement (1) the relevant context words around the word "*class*" are "enjoy", "super", "interesting", and in statement (2) the relevant context words around the word "*class*" are "boring", "asleep." The term "*class*" thus gains different context words in different communicative domains. In short, because the student's word choice when speaking to the Professor is different from the word choice when speaking to their

[226] Hammond, Kenneth R. "Coherence and correspondence theories in judgment and decision making." (2000).
[227] Hammond, Kenneth R. *Human judgment and social policy: Irreducible uncertainty, inevitable error, unavoidable injustice*. Oxford University Press, USA, 1996.
[228] Blair, J. Pete, Torsten O. Reimer, and Timothy R. Levine. "The role of consistency in detecting deception: The superiority of correspondence over coherence." *Communication Studies* 69, no. 5 (2018): 483-498.

friends, we can claim, based solely on lexical data, that the student is dishonest due to their incoherence and inconsistency.

Extending this to the present research question, it is possible that the aforementioned coherence competence approach to the investigation of dishonesty can be applied to the question of corporate honesty on ESG. This can be done by examining contextual word use surrounding ESG and other related terms as generated by corporate actors in candid versus official corporate discourse channels. For instance, if the official discourse channel contains the phrase "ESG is great" while the candid discourse channel contains the phrase "ESG is a sham," then we can use this information to establish a case for of corporate dishonesty on the issue of ESG. Thus, incoherence in word use on the subject of ESG across distinct corporate discourse channels can be argued to be indicative of dishonesty. From the perspective of word embedding modeling, which will be discussed later, different context words in different domains would essentially signify a change in meaning, which can be quantified using lexical semantic change techniques proposed in Hamilton et al.'s and Schlechtweg et al.'s work.[229]

Before proceeding further, it is worth briefly mentioning that Hammond's coherence criteria is not the only method for the evaluation of truth statements, and alternatively, Hammond discusses the correspondence theory, which "evaluates the correspondence between the judgment and the empirical fact that is the object of the judgment."[230] Thus, the distinction between coherence and correspondence theories of truth is that correspondence emphasizes the external empirical facts that can be verified, whereas coherence deals with consistency of statements, facts and beliefs, regardless of whether such facts can be verified.[231] Blair et al. provide a useful example of this distinction, noting that "the view that the world was flat was once a dominant and widely shared belief, which was part of a coherent belief system."[232] Thus, in the words of Hammond, unlike the coherence theory, "the correspondence theory asks not for logic or consistency but only for accuracy – the correspondence of judgments with the facts."[233] It is also worth pointing out that Hammond's coherence-correspondence distinction is not without its critics, and furthermore, is not the only theory on truth, judgement and

---

[229] Hamilton, William L., Jure Leskovec, and Dan Jurafsky. "Diachronic word embeddings reveal statistical laws of semantic change." *arXiv preprint arXiv:1605.09096* (2016).
Schlechtweg, Dominik, Anna Hätty, Marco Del Tredici, and Sabine Schulte im Walde. "A wind of change: Detecting and evaluating lexical semantic change across times and domains." *arXiv preprint arXiv:1906.02979* (2019).
[230] Hammond, Kenneth R. *Beyond rationality: The search for wisdom in a troubled time*. Oxford University Press, 2007.
[231] *Ibid*
[232] Blair, J. Pete, Torsten O. Reimer, and Timothy R. Levine. "The role of consistency in detecting deception: The superiority of correspondence over coherence." *Communication Studies* 69, no. 5 (2018): 483-498.
[233] Hammond, Kenneth R. *Human judgment and social policy: Irreducible uncertainty, inevitable error, unavoidable injustice*. Oxford University Press, USA, 1996. Emphasis in the original work.

decision-making.[234] For present purposes, we simply adopt the coherence view as a means of investigating the question of honesty in corporate discourse. An example of a correspondence-based approach to corporate honesty on the issue of ESG is represented by the Raghanundan and Rajgobal study, which compares the purported commitment to social responsibility, with facts, namely that socially responsible firms pollute more.[235]

2.5.1 Synchronic Lexical Semantic Change Detection in Distinct Corporate Discourse Channels

To compare the meaning of ESG related terms across different corporate discourse channels we first begin by preprocessing our corpus, which includes lemmatization, url removal, stopword removal, etc. We split our dataset into two distinct discourse channels, namely the candid corporate discourse channel, which contains textual data from earning conference call transcripts, and the official corporate discourse channel, which contains textual data from official corporate reports, namely the mandatory 10-Ks and 10-Qs supplemented with textual data from voluntary ESG reports.

Our next step is to establish a dictionary of words for comparison across the two channels. To do this, we construct two topic models separately on both the candid and official corporate discourse channels using Latent Dirichlet Allocation (LDA) topic modeling algorithm.[236] We set the number of topics at a fixed number for both discourse channels. For future research, we intend to test and find an optimal number of topics using various proposed methods and measurements, from topic coherence to minimization of perplexity.[237] After extracting a list of topics and topic descriptor words for each discourse channel, we use the Jaccard similarity metric to measure the top

---

[234] Polonioli, Andrea. "The uses and abuses of the coherence–correspondence distinction." *Frontiers in Psychology* 6 (2015): 507.

[235] Raghunandan, Aneesh, and Shivaram Rajgopal. "Do socially responsible firms walk the talk?." *Available at SSRN 3609056* (2022).

[236] Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent dirichlet allocation." *Journal of machine Learning research* 3, no. Jan (2003)

[237] Zhao, Weizhong, James J. Chen, Roger Perkins, Zhichao Liu, Weigong Ge, Yijun Ding, and Wen Zou. "A heuristic approach to determine an appropriate number of topics in topic modeling." In *BMC bioinformatics*, vol. 16, pp. 1-10. BioMed Central, 2015.
Arun, Rajkumar, Venkatasubramaniyan Suresh, C. E. Veni Madhavan, and M. N. Narasimha Murthy. "On finding the natural number of topics with latent dirichlet allocation: Some observations." In *Advances in Knowledge Discovery and Data Mining: 14th Pacific-Asia Conference, PAKDD 2010, Hyderabad, India, June 21-24, 2010. Proceedings. Part I 14*, pp. 391-402. Springer Berlin Heidelberg, 2010.
See also: Greene, Derek, Derek O'Callaghan, and Pádraig Cunningham. "How many topics? stability analysis for topic models." In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2014, Nancy, France, September 15-19, 2014. Proceedings, Part I 14*, pp. 498-513. Springer Berlin Heidelberg, 2014.

descriptor words of topics across the two discourse channels.[238] The Jaccard similarity measure is bounded between 0 and 1, and thus, when two topics across two discourse channels have identical topic descriptor words, the Jaccard similarity will be 1, and conversely, if all the words are different, the Jaccard similarity will be 0.[239] At this point, we closely read the topics which are similar across the two corporate discourse channels, i.e. above a certain Jaccard similarity threshold, and we interpret the words from these topics as a representation of cross-corporate issues, such as for instance, tax, ESG, supply chain issues etc. Non-ESG topics are necessary for comparison as they establish a baseline for subsequent lexical semantic change calculations.

After developing our dictionary of comparison words, we build two 100-dimensional word2vec skip-gram models for each corporate discourse channel.[240] Following Hamilton et al.[241] and Schlechtweg et al.[242] we align the word2vec embeddings using the orthogonal Procrustes alignment technique. This step ensures the comparability of word vector representations across the models for lexical semantic change detection purposes. Using the dictionary of words that represent cross-corporate issues that were established via LDA and Jaccard similarity methods, we subsequently examine the cosine distances that these words have across the two corporate discourse channels. Cosine distance for an individual word across two corporate discourse channels, or domains, is defined as:

$$Cosine\ Distance\left(word_{candid}, word_{official}\right) := 1 - Cosine(word_{candid}, word_{official})$$

Cosine distance for an individual word across two word embedding models can be interpreted "as a measure of semantic change and displacement."[243] Since cosine distance is bounded between 0 and 2, a cosine distance of 0 signifies that there is no displacement and thus no lexical semantic change for an individual word between the candid and the official word embedding models. Conversely, a cosine distance of 2 signifies that there is maximum displacement of a word between the candid and the official word embedding models, which can be interpreted as a complete change in context words and thus indicative of considerable semantic change. However, a non-zero cosine distance score

[238] Mantyla, Mika V., Maelick Claes, and Umar Farooq. "Measuring LDA topic stability from clusters of replicated runs." In *Proceedings of the 12th ACM/IEEE international symposium on empirical software engineering and measurement*, pp. 1-4. 2018.
[239] *Ibid.*
[240] Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. "Distributed representations of words and phrases and their compositionality." *Advances in neural information processing systems* 26 (2013).
[241] Hamilton, William L., Jure Leskovec, and Dan Jurafsky. "Diachronic word embeddings reveal statistical laws of semantic change." *arXiv preprint arXiv:1605.09096* (2016).
[242] Schlechtweg, Dominik, Anna Hätty, Marco Del Tredici, and Sabine Schulte im Walde. "A wind of change: Detecting and evaluating lexical semantic change across times and domains." *arXiv preprint arXiv:1906.02979* (2019).
[243] Hamilton, William L., Jure Leskovec, and Dan Jurafsky. "Diachronic word embeddings reveal statistical laws of semantic change." *arXiv preprint arXiv:1605.09096* (2016).

by itself does not necessarily imply lexical semantic change, as it could simply be a consequence of the written versus spoken nature of the textual data in the two discourse channels. Therefore, it is important to also examine the cosine distances for other cross-corporate non-ESG issues like for instance "taxation" as they can provide a baseline value for the lexical semantic divergence scores. Accordingly, if ESG words, which represent the issue of ESG, have a significantly higher cosine distance scores than tax words, then it can be argued that that ESG words, and thus the ESG issue, is experiencing a lexical semantic change between two discourse channels when compared to non-ESG issues like taxation. Relative similarity in cosine distances between ESG and non-ESG issues would alternatively imply a lack of lexical semantic change across the two corporate discourse channels. Thus, it could be stated that higher cosine distance scores for ESG words when compared to non-ESG words would imply inconsistency and thus corporate dishonesty on the issue of ESG.

2.6 Results and Discussion

Our results in Figure 1 below indicate that ESG topic descriptor words, in black, do not exhibit a significant difference in cosine distances across corporate discourse channels when compared to other cross-corporate topic descriptor words, for instance, "corporate voting" in gray. The results seem to suggest that prima facie, corporate actors use similar lexicon and context words when they talk about the issue of ESG in both candid and official discourse channels which explains the lack of detectable lexical semantic change across these two domains.

There can be multiple interpretations of the results we observe. Firstly, an optimistic interpretation of the evidence can confirm the story of consistency on the issue of ESG across distinct corporate discourse channels, and thus, it could be argued that corporate actors seem to be honest, or lexically consistent, when it comes to their commitments to ESG. However, a pessimistic and cynical interpretation of the results is also possible. Given the fact that language management is an important part of corporate communication and discourse, and thus, consistency plays an important role in corporate communication, perhaps it is not surprising that corporate actors are so coherent on the issue of ESG. Corporate managers are sophisticated actors and expecting a CEO of a major corporation to say their true opinion on ESG at an earnings conference call, for example by stating that ESG is just a marketing ploy, is likely an impossible expectation, especially given that these earning calls are recorded. As stated earlier, corporate actors have a strong incentive to be incredibly careful with their statements, [244] which is the reason we do not see a significant lexical semantic divergence on the ESG topic, or any

[244] See also: Bargiela, Francesca, and Sandra J. Harris. *Managing language: The discourse of corporate meetings*. Vol. 44. John Benjamins Publishing, 1997.
See also: Amernic, Joel, and Russell Craig. CEO-speak: The language of corporate leadership. McGill-Queen's Press-MQUP, 2006.

other topic for that matter. This also seems to confirm Li's insight that even in "different discourse channels, there are likely influences of similar economic factors."[245] The interpretation that sophisticated actors can generate and manage their language according to what's expected of them is of course not novel.[246] In fact, their ability to take great care of their language is probably the reason why they were hired as corporate officers in the first place. This is perhaps the fundamental limitation of separating corporate communication into candid and official discourse channels – perhaps corporate communication is identical regardless of the discourse channel that is used for communication because corporate communication must be, in essence, organizationally coherent.[247]

There are also other limitations to our approach as well. For instance, the cross-corporate dictionary of ESG words that we obtained using our topic modeling approach seem to be focused more on the Environmental, or "E" aspect of ESG, and thus lack social, governance or other stakeholder terms. This likely reflects a greater amount of discussion on the issues of the environment, which are in turn based on regulatory and theoretical foundations underpinning environmental regulations, rooted in matters concerning carbon accounting and the Kyoto Protocol.[248] Furthermore, such corporate discussions are also likely a reflection of the weight given to issues of climate change in societal discourse overall.[249] Interestingly, the focus on the environment also seems to confirm Jaworska's insight that discussions on climate change in corporate discourse have shifted to being seen as a "risk" for business, and thus, according to Jaworska, "the increased attention to risks in recent years marks a discursive shift, which turns climate

---

[245] Li, Feng. "Textual analysis of corporate disclosures: A survey of the literature." *Journal of accounting literature* 29 (2010): 143.

[246] Bargiela, Francesca, and Sandra J. Harris. *Managing language: The discourse of corporate meetings*. Vol. 44. John Benjamins Publishing, 1997.
Amernic, Joel, and Russell Craig. *CEO-speak: The language of corporate leadership*. McGill-Queen's Press-MQUP, 2006.

[247] Bargiela, Francesca, and Sandra J. Harris. *Managing language: The discourse of corporate meetings*. Vol. 44. John Benjamins Publishing, 1997.

[248] Stechemesser, Kristin, and Edeltraud Guenther. "Carbon accounting: a systematic literature review." *Journal of Cleaner Production* 36 (2012): 17-38.
See also: Csutora, Maria, and Gabor Harangozo. "Twenty years of carbon accounting and auditing–a review and outlook." *Society and Economy* 39, no. 4 (2017): 459-480.
For economic-theoretic discussions see: Chichilnisky, Graciela, and Geoffrey Heal. "Who should abate carbon emissions?: An international viewpoint." *Economics Letters* 44, no. 4 (1994): 443-449.
See also: Bushnell, James B. "The economics of carbon offsets." In *The design and implementation of US climate policy*, pp. 197-209. University of Chicago Press, 2011.

[249] Dechezleprêtre, Antoine, Adrien Fabre, Tobias Kruse, Bluebery Planterose, Ana Sanchez Chico, and Stefanie Stantcheva. *Fighting climate change: International attitudes toward climate policies*. No. w30265. National Bureau of Economic Research, 2022.

change from an agentless object into an agentive subject."[250] Another limitation stems from the use of Jaccard similarity metric, which only captures similarity of lexical use via the comparison of topic descriptor words for topics across corporate discourse channels, which does not necessarily signify identical topics being discussed. In other words, simply comparing topic descriptor words might not capture the nuance of the corporate discussions on these issues. Overall, an alternative method for constructing a dictionary of ESG and other cross-corporate issues is thus necessary, perhaps by utilizing novel large language model-based approaches for extracting information from financial texts, such as the recently developed FinBERT.[251]

Nevertheless, despite all these limitations, the present research proposes a novel approach to the question of exploring corporate honesty on the issue of ESG by adopting a word embedding based synchronic lexical semantic change detection approach to two proposed corporate discourse channels. Future research avenues seem promising however, with novel data in this space being made available, such as the recently compiled and released data of recordings of earning calls,[252] and furthermore, with other methods, such as Ferrari's word injection approach[253] and the more recent FrameAxis approach for measuring custom frames and biases with word embeddings,[254] and additionally, the application of cultural dimension of word embeddings-based approaches to this question,[255] and finally, the application of large language model-based FinBERT.[256]

---

[250] Jaworska, Sylvia. "Change but no climate change: Discourses of climate change in corporate social responsibility reporting in the oil industry." *International Journal of Business Communication* 55, no. 2 (2018): 194-219.
[251] Huang, Allen H., Hui Wang, and Yi Yang. "FinBERT: A large language model for extracting information from financial text." *Contemporary Accounting Research* 40, no. 2 (2023): 806-841.
[252] Li, Jiazheng, Linyi Yang, Barry Smyth, and Ruihai Dong. "Maec: A multimodal aligned earnings conference call dataset for financial risk prediction." In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pp. 3063-3070. 2020.
See also: Qin, Yu, and Yi Yang. "What you say and how you say it matters: Predicting stock volatility using verbal and vocal cues." In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 390-401. 2019.
[253] Ferrari, Alessio, Beatrice Donati, and Stefania Gnesi. "Detecting domain-specific ambiguities: an NLP approach based on Wikipedia crawling and word embeddings." In *2017 IEEE 25th International Requirements Engineering Conference Workshops (REW)*, pp. 393-399. IEEE, 2017.
[254] Kwak, Haewoon, Jisun An, Elise Jing, and Yong-Yeol Ahn. "FrameAxis: characterizing microframe bias and intensity with word embedding." *PeerJ Computer Science* 7 (2021): e644.
[255] Kozlowski, Austin C., Matt Taddy, and James A. Evans. "The geometry of culture: Analyzing the meanings of class through word embeddings." *American Sociological Review* 84, no. 5 (2019): 905-949.
[256] Huang, Allen H., Hui Wang, and Yi Yang. "FinBERT: A large language model for extracting information from financial text." *Contemporary Accounting Research* 40, no. 2 (2023): 806-841.

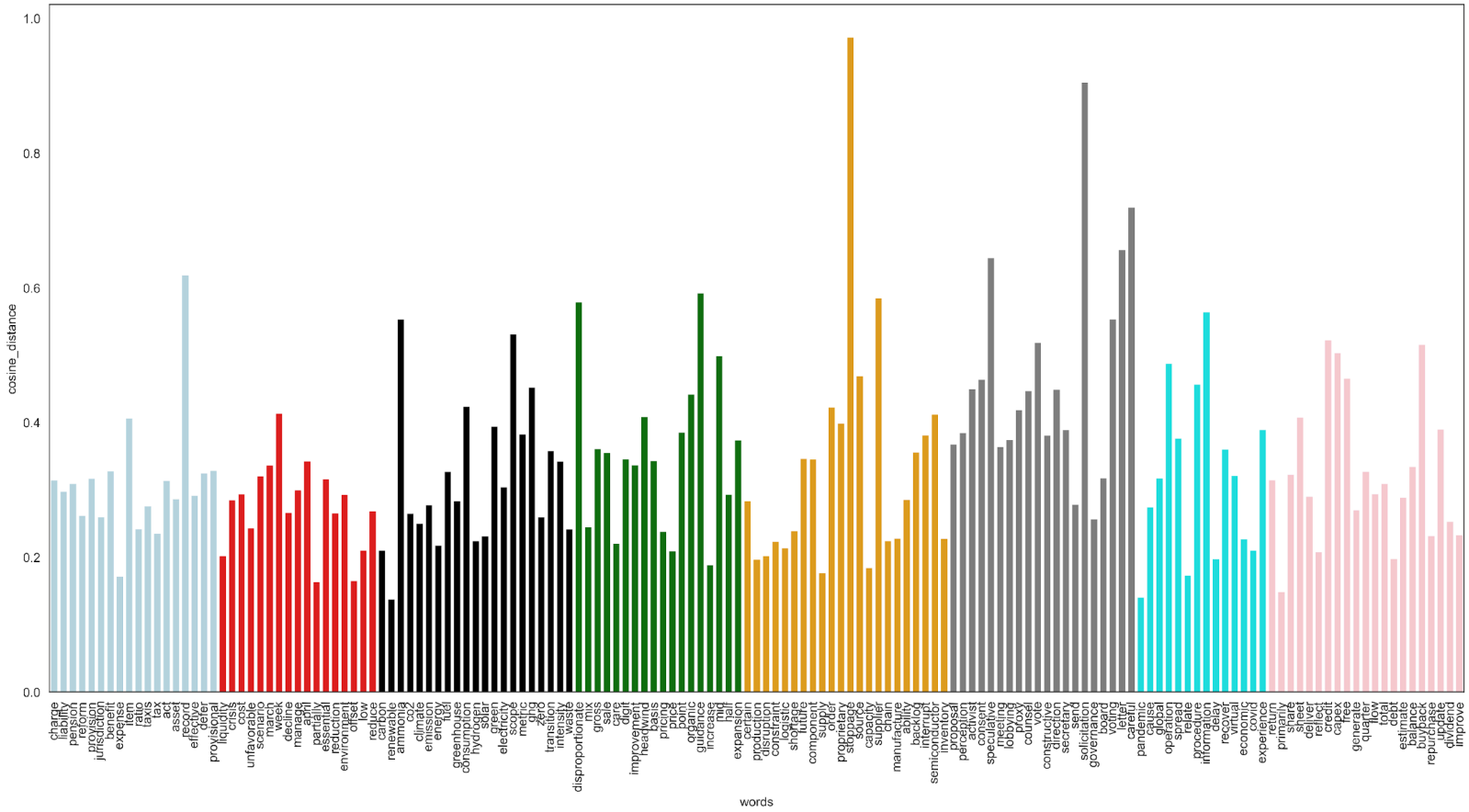

Figure 1: Cosine distances for cross-corporate issues. The X-axis represents individual words color-coded by cross-corporate issues. ESG-related words are colored black. For cosine distance comparison, other cross-corporate issues are also included. For instance, words relating to issues of corporate voting are in gray. The words on the X-axis are extracted using LDA topics and Jaccard similarity of topic descriptor words across discourse channels. The Y-axis represents the cosine distances between an individual word in candid versus official discourse channels. Values are bounded between 0 and 2, with 0 signifying no semantic displacement and 2 signifying complete semantic displacement.

2.7 Conclusion

In the present work, we proposed a novel methodology for exploring the question of corporate honesty on the issue of ESG. To tackle this question, we adopted lexical semantic change detection approaches to quantify change in meaning of ESG related terms across two distinct corporate discourse channels, namely the candid and the official discourse channels. Our results suggest that corporate communication is indeed lexically consistent on the issue of ESG across these corporate discourse channels, which, naively and at a first glance, could be interpreted as honesty on the issue of ESG. Nevertheless, given the care that sophisticated corporate actors exercise when communicating, especially on such a novel issue as ESG, it is perhaps unsurprising that corporate actors are so consistent on this issue. Thus, the semantic similarity of ESG words across discourse channels could be interpreted as a reflection of the underlying sophisticated nature of corporate actors and executives, rather than their honesty. This again confirms Li's pertinent and accurate insight that even in "different discourse channels, there are likely influences of similar economic factors." [257] Furthermore, an artificial separation of corporate communication into separate discourse channels needs further investigation, and perhaps, these communication channels are not so distinct after all. Further work is necessary to explore this question, with promising avenues being offered by novel methods and data sources.

[257] Li, Feng. "Textual analysis of corporate disclosures: A survey of the literature." *Journal of accounting literature* 29 (2010): 143.

2.8 References

Amernic, Joel, and Russell Craig. *CEO-speak: The language of corporate leadership*. McGill-Queen's Press-MQUP, 2006.

Arun, Rajkumar, Venkatasubramaniyan Suresh, C. E. Veni Madhavan, and M. N. Narasimha Murthy. "On finding the natural number of topics with latent dirichlet allocation: Some observations." In *Advances in Knowledge Discovery and Data Mining: 14th Pacific-Asia Conference, PAKDD 2010, Hyderabad, India, June 21-24, 2010. Proceedings. Part I 14*, pp. 391-402. Springer Berlin Heidelberg, 2010.

Bargiela, Francesca, and Sandra J. Harris. *Managing language: The discourse of corporate meetings*. Vol. 44. John Benjamins Publishing, 1997.

Berk, Jonathan, and Jules H. Van Binsbergen. "The impact of impact investing." (2021).

Blair, J. Pete, Torsten O. Reimer, and Timothy R. Levine. "The role of consistency in detecting deception: The superiority of correspondence over coherence." *Communication Studies* 69, no. 5 (2018): 483-498.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent dirichlet allocation." *Journal of machine Learning research* 3, no. Jan (2003)

Boehm, Paul D., Gregory S. Douglas, William A. Burns, Paul J. Mankiewicz, David S. Page, and A. Edward Bence. "Application of petroleum hydrocarbon chemical fingerprinting and allocation techniques after the Exxon Valdez oil spill." *Marine Pollution Bulletin* 34, no. 8 (1997): 599-613.

Brown, Stephen V., and Jennifer Wu Tucker. "Large‐sample evidence on firms' year‐over‐year MD&A modifications." *Journal of Accounting Research* 49, no. 2 (2011): 309-346.

Bushnell, James B. "The economics of carbon offsets." In *The design and implementation of US climate policy*, pp. 197-209. University of Chicago Press, 2011.

Camilleri, Mark Anthony. "Theoretical insights on integrated reporting: The inclusion of non-financial capitals in corporate disclosures." *Corporate Communications: An International Journal* 23, no. 4 (2018): 567-581.

Campbell, John L., Hsinchun Chen, Dan S. Dhaliwal, Hsin-min Lu, and Logan B. Steele. "The information content of mandatory risk factor disclosures in corporate filings." *Review of Accounting Studies* 19 (2014): 396-455.

Cazier, Richard A., and Ray J. Pfeiffer. "Why are 10-K filings so long?." *Accounting Horizons* 30, no. 1 (2016): 1-21.

Chen, Shuping, Stephan Hollander, and Kelvin Law. "In search of interaction." *Available at SSRN 2449341* (2016).

Chichilnisky, Graciela, and Geoffrey Heal. "Who should abate carbon emissions?: An international viewpoint." *Economics Letters* 44, no. 4 (1994): 443-449.

Chiu, Tiffany, Feiqi Huang, Yue Liu, and Miklos A. Vasarhelyi. "The impact of non-timely 10-Q filings and audit firm size on audit fees." *Managerial Auditing Journal* 33, no. 5 (2018): 503-516.

Christensen, Dane M., George Serafeim, and Anywhere Sikochi. "Why is corporate virtue in the eye of the beholder? The case of ESG ratings." *The Accounting Review* 97, no. 1 (2022): 147-175.

Csutora, Maria, and Gabor Harangozo. "Twenty years of carbon accounting and auditing–a review and outlook." *Society and Economy* 39, no. 4 (2017): 459-480.

Damodaran, Aswath. *Corporate finance*. Hoboken: Wiley, 1996.

Danescu-Niculescu-Mizil, Cristian, Lillian Lee, Bo Pang, and Jon Kleinberg. "Echoes of power: Language effects and power differences in social interaction." In *Proceedings of the 21st international conference on World Wide Web*, pp. 699-708. 2012.

Dechezleprêtre, Antoine, Adrien Fabre, Tobias Kruse, Bluebery Planterose, Ana Sanchez Chico, and Stefanie Stantcheva. *Fighting climate change: International attitudes toward climate policies*. No. w30265. National Bureau of Economic Research, 2022.

Delmas, Magali A., and Vanessa Cuerel Burbano. "The drivers of greenwashing." *California management review* 54, no. 1 (2011): 64-87.

Dorfleitner, Gregor, Gerhard Halbritter, and Mai Nguyen. "Measuring the level and risk of corporate responsibility–An empirical comparison of different ESG rating approaches." Journal of Asset Management 16 (2015): 450-466.

Feely, Alan J., and Anne‐Wil Harzing. "Language management in multinational companies." *Cross Cultural Management: an international journal* 10, no. 2 (2003): 37-52.

Ferrari, Alessio, Beatrice Donati, and Stefania Gnesi. "Detecting domain-specific ambiguities: an NLP approach based on Wikipedia crawling and word embeddings." In *2017 IEEE 25th International Requirements Engineering Conference Workshops (REW)*, pp. 393-399. IEEE, 2017.

Galasinski, Dariusz. *The language of deception: A discourse analytical study*. Sage Publications, 2000.

Graciarena, Martin, Elizabeth Shriberg, Andreas Stolcke, Frank Enos, Julia Hirschberg, and Sachin Kajarekar. "Combining prosodic lexical and cepstral systems for deceptive speech detection." In 2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings, vol. 1, pp. I-I. IEEE, 2006.

Greene, Derek, Derek O'Callaghan, and Pádraig Cunningham. "How many topics? stability analysis for topic models." In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2014, Nancy, France, September 15-19, 2014. Proceedings, Part I 14*, pp. 498-513. Springer Berlin Heidelberg, 2014.

Gregory Jr, Stanford W., and Stephen Webster. "A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perceptions." *Journal of personality and social psychology* 70, no. 6 (1996): 1231.

Hamilton, William L., Jure Leskovec, and Dan Jurafsky. "Diachronic word embeddings reveal statistical laws of semantic change." *arXiv preprint arXiv:1605.09096* (2016).

Hammond, Kenneth R. "Coherence and correspondence theories in judgment and decision making." (2000).

Huang, Allen H., Hui Wang, and Yi Yang. "FinBERT: A large language model for extracting information from financial text." *Contemporary Accounting Research* 40, no. 2 (2023): 806-841.

Jaworska, Sylvia, and Anupam Nanda. "Doing well by talking good: A topic modelling-assisted discourse study of corporate social responsibility." *Applied Linguistics* 39, no. 3 (2018): 373-399.

Jaworska, Sylvia. "Change but no climate change: Discourses of climate change in corporate social responsibility reporting in the oil industry." *International Journal of Business Communication* 55, no. 2 (2018): 194-219.

Jaworska, Sylvia. "Corporate discourse." (2020).

Kim, Hyogon, Eunmi Lee, and Donghee Yoo. "Do SEC filings indicate any trends? Evidence from the sentiment distribution of forms 10-K and 10-Q with FinBERT." *Data Technologies and Applications* ahead-of-print (2023).

Knoepfel, Ivo. "Who Cares Wins: Connecting Financial Markets to a Changing World, UN Environment Programme." (2004).

Kozlowski, Austin C., Matt Taddy, and James A. Evans. "The geometry of culture: Analyzing the meanings of class through word embeddings." *American Sociological Review* 84, no. 5 (2019): 905-949.

Kwak, Haewoon, Jisun An, Elise Jing, and Yong-Yeol Ahn. "FrameAxis: characterizing microframe bias and intensity with word embedding." *PeerJ Computer Science* 7 (2021): e644.

Laine, Matias. "Meanings of the term ‘sustainable development’ in Finnish corporate disclosures." In *Accounting Forum*, vol. 29, no. 4, pp. 395-413.

Landi, Giovanni, and Mauro Sciarelli. "Towards a more ethical market: the impact of ESG rating on corporate financial performance." *Social responsibility journal* 15, no. 1 (2018): 11-27.

Lewis, Craig, and Steven Young. "Fad or future? Automated analysis of financial text and its implications for corporate reporting." *Accounting and Business Research* 49, no. 5 (2019): 587-615.

Li, Feng. "Textual analysis of corporate disclosures: A survey of the literature." *Journal of accounting literature* 29 (2010): 143.

Li, Jiazheng, Linyi Yang, Barry Smyth, and Ruihai Dong. "Maec: A multimodal aligned earnings conference call dataset for financial risk prediction." In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pp. 3063-3070. 2020.

Li, Ting-Ting, Kai Wang, Toshiyuki Sueyoshi, and Derek D. Wang. "ESG: Research progress and future prospects." *Sustainability* 13, no. 21 (2021): 11663.

Lund, Dorothy S., and Elizabeth Pollman. "The corporate governance machine." *Colum. L. Rev.* 121 (2021): 2563.

Mantyla, Mika V., Maelick Claes, and Umar Farooq. "Measuring LDA topic stability from clusters of replicated runs." In *Proceedings of the 12th ACM/IEEE international symposium on empirical software engineering and measurement*, pp. 1-4. 2018.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. "Distributed representations of words and phrases and their compositionality." *Advances in neural information processing systems* 26 (2013).

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. "Efficient estimation of word representations in vector space." *arXiv preprint arXiv:1301.3781* (2013).

Mulligan, Thomas. "A critique of Milton Friedman's essay ‘the social responsibility of business is to increase its profits’." *Journal of Business Ethics* 5 (1986): 265-269.

Pérez, Lucy, Vivian Hunt, Hamid Samandari, Robin Nuttall, and Krysta Biniek. "Does ESG really matter—and why." *McKinsey Quarterly* (2022).

Polonioli, Andrea. "The uses and abuses of the coherence–correspondence distinction." *Frontiers in Psychology* 6 (2015): 507.

Qin, Yu, and Yi Yang. "What you say and how you say it matters: Predicting stock volatility using verbal and vocal cues." In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 390-401. 2019.

Raghunandan, Aneesh, and Shivaram Rajgopal. "Do socially responsible firms walk the talk?." *Available at SSRN 3609056* (2022).

Rock, Edward B. "For Whom is the Corporation Managed in 2020?: The Debate over Corporate Purpose." European Corporate Governance Institute-Law Working Paper 515 (2020): 20-16.

Schlechtweg, Dominik, Anna Hätty, Marco Del Tredici, and Sabine Schulte im Walde. "A wind of change: Detecting and evaluating lexical semantic change across times and domains." *arXiv preprint arXiv:1906.02979* (2019).

Schwab, Klaus. *Stakeholder capitalism: A global economy that works for progress, people and planet*. John Wiley & Sons, 2021.

Seneca, Lucius Annaeus, *Seneca: Moral and political essays*. Cambridge University Press, 1995.

Sprinkle, Geoffrey B., and Laureen A. Maines. "The benefits and costs of corporate social responsibility." *Business Horizons* 53, no. 5 (2010): 445.

Stechemesser, Kristin, and Edeltraud Guenther. "Carbon accounting: a systematic literature review." *Journal of Cleaner Production* 36 (2012): 17-38.

Strine Jr, Leo E. "The dangers of denial: The need for a clear-eyed understanding of the power and accountability structure established by the Delaware general corporation law." *Wake Forest L. Rev.* 50 (2015): 761.

Strine Jr, Leo E., Justin L Brooke, Kyle M. Diamond, and Derrick L. Parker Jr. "It's Time to Focus on the “G” in ESG." *Harvard Business Review, November* 18 (2022).

Suchman, Mark C. "Managing legitimacy: Strategic and institutional approaches." Academy of management review 20, no. 3 (1995): 571-610.

Voigt, Rob, Nicholas P. Camp, Vinodkumar Prabhakaran, William L. Hamilton, Rebecca C. Hetey, Camilla M. Griffiths, David Jurgens, Dan Jurafsky, and Jennifer L. Eberhardt. "Language from police body camera footage shows racial disparities in officer respect." *Proceedings of the National Academy of Sciences* 114, no. 25 (2017): 6521-6526.

Weinstein, James. "The corporate ideal in the liberal state: 1900-1918." (1969).

Whelan, Tenise, U. Atz, T. Van Holt, and C. Clark. "Uncovering the Relationship by Aggregating Evidence from 1,000 Plus Studies Published between 2015–2020." 2022.

Widyawati, Luluk. "A systematic literature review of socially responsible investment and environmental social governance metrics." *Business Strategy and the Environment* 29, no. 2 (2020): 619-637.

Xia, Yuqiang, and Michel C. Boufadel. "Lessons from the Exxon Valdez oil spill disaster in Alaska." *Disaster Advances* 3, no. 4 (2010): 270-273.

Yu, Ellen Pei‐yi, Christine Qian Guo, and Bac Van Luu. "Environmental, social and governance transparency and firm value." *Business Strategy and the Environment* 27, no. 7 (2018): 987-1004.

Zhao, Weizhong, James J. Chen, Roger Perkins, Zhichao Liu, Weigong Ge, Yijun Ding, and Wen Zou. "A heuristic approach to determine an appropriate number of topics in topic modeling." In *BMC bioinformatics*, vol. 16, pp. 1-10. BioMed Central, 2015.

Žižek, Slavoj. *First as tragedy, then as farce*. Verso, 2009.

3 Pre-Text: Epistemological Perspectives on Legal Reasoning, Natural Language Processing and Artificial Intelligence

Abstract

With the rise of legal AI and the recent use of ChatGPT for judicial decision-making, this essay investigates the epistemological questions surrounding the possibilities of legal decision-making being done with the help of AI. We begin by discussing legal reasoning, focusing on Jean-Michel Berthelot's schemas of intelligibility-based theory of epistemology. We argue that legal reasoning and thus legal textual data essentially reflect these schemas of intelligibility. We proceed with a discussion of large language models, particularly focusing on their propensity to capture the various biases in the training data. Putting these arguments together, we predict that, while large language models are not explicitly designed to capture epistemological biases, they will still exhibit latent legal epistemological biases rooted in schemas of intelligibility-based patterns of legal reasoning as observed in the text. This in turn necessitates a re-examination of legal textual data with a focus on explicitly categorizing these legal epistemological biases and studying the linguistic cues associated with these schemes of intelligibility.

Jamais deux hommes ne jugèrent pareillement de même chose,
et est impossible de voir deux opinions semblables exactement,
non seulement en divers hommes, mais en même homme à diverses heures.

[No two men ever judged alike of the same thing,
and it is impossible to find two opinions exactly similar,
not only in different men but in the same man at different times.]

Michel de Montaigne, *On Experience* (1580)[258]

[258] de Montaigne, Michel. *Essays* (1595), translated by JM Cohen, Penguin Books. (1957).

3.1 Introduction

In recent years, advances in machine learning, natural language processing and artificial intelligence have undoubtedly had a transformative effect on many aspects of our lives.[259] While these advances bring tremendous benefits, their quick rise in various domains raises important socio-economic, socio-political, and socio-legal questions.[260]

Perhaps the question of AI's impact on labor is currently the most sensitive one as many industries and professions are currently being disrupted by these technologies.[261] For instance, a recent Goldman Sachs report by Briggs and Kodnani titled "The Potentially Large Effects of Artificial Intelligence on Economic Growth" argues that "roughly two-thirds of current jobs are exposed to some degree of AI automation, and that generative AI could substitute up to one-fourth of current work."[262] However, most surprising are their findings that the second highest share of industry employment "exposed to automation" is "legal services", with 44% of legal work predicted to be automated by generative AI.[263] Legal services were outmatched only by "Office and Administrative Support", with 46% of work predicted to be automated in this space.[264] Beyond such macroeconomic predictions, currently, the ongoing SAG-AFTRA and Writers Guild of America (WGA) labor strikes in Hollywood show that large language model based generative AI's impacts are already being felt. For instance, John August, a screenwriter for "Charlie's Angels," stated that he wants WGA to make sure that "ChatGPT and its cousins can't be credited with writing a screenplay."[265]

The rise of large language model-based AI is indeed occurring in the legal industry as well,[266] supporting Briggs and Kodnani's predictions. For instance, some lawyers are

---

[259] See for example: Vamathevan, Jessica, Dominic Clark, Paul Czodrowski, Ian Dunham, Edgardo Ferran, George Lee, Bin Li et al. "Applications of machine learning in drug discovery and development." *Nature reviews Drug discovery* 18, no. 6 (2019): 463-477.
See also: Portugal, Ivens, Paulo Alencar, and Donald Cowan. "The use of machine learning algorithms in recommender systems: A systematic review." *Expert Systems with Applications* 97 (2018): 205-227.

[260] Arrieta, Alejandro Barredo, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García et al. "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI." *Information fusion* 58 (2020): 82-115.

[261] Reinert, Hugo, and Erik S. Reinert. "Creative destruction in economics: Nietzsche, Sombart, schumpeter." *Friedrich Nietzsche (1844–1900) Economy and Society* (2006): 55-85.

[262] Hatzius, Jan. "The Potentially Large Effects of Artificial Intelligence on Economic Growth (Briggs/Kodnani)." Goldman Sachs (2023).

[263] *Ibid.*

[264] *Ibid.*

[265] Klippenstein, Ken. "As Actors Strike for AI Protections, Netflix Lists $900,000 AI Job." The Intercept, July 25, 2023. https://theintercept.com/2023/07/25/strike-hollywood-ai-disney-netflix/.

[266] Coyer, Cassandre. "A Welcome Boon or a Dreaded Replacement? Paralegals Split on Generative AI." Law.com, July 25, 2023. Available at:

already using these large language models in their work, noting the simplicity of the process: "load in a case's documents and ask the software to draft deposition questions, for example, and in a few minutes, it will spit out a list of pertinent questions."[267] Judging by this, Legal AI in all its manifestations is thus a possible eventuality.

Putting aside the longstanding questions regarding technology's relationship with labor,[268] this present work seeks to explore the epistemological implications of the possible scenario where large language model-based AI technology, like for instance ChatGPT, is used beyond the drafting of contracts or the generation of deposition questions, but rather, is used for legal reasoning and legal decision-making as such.

In the first part of this essay, I will be motivating the study with a discussion of the case of Judge Juan Manuel Padilla who was the first judge to use ChatGPT to render a legal decision in a case. In the second part, I will discuss Jean-Michelle Berthelot schema of intelligibility-based framework, as explained and expounded upon in Professor Samuel's work on legal epistemology, arguing that this theory captures the essential aspects of legal reasoning and legal decision-making.[269] For instance, Berthelot's model is able to explain one of the central questions of legal theory, which is perhaps most noticeable in so-called "hard cases", namely, how a single factual and legal situation can in fact often result in multiple legal opinions. I will be linking Berthelot's theory to legal textual data, showing that legal text essentially mirrors the schema of intelligibility-based reasoning patterns employed by jurists. In the third part, I will be discussing large language models. Of particular interest will be the question of large language models' propensity towards capturing bias and their "garbage in, garbage out" reflection of the underlying data. Furthermore, I will discuss the possible capabilities of large language models to capture patterns of reasoning in the text. Finally, I will tie the two arguments together, namely, that large language models and AI capture the bias in the underlying training data, and that legal reasoning involves the application of schemes of intelligibility as predicted by Berthelot's model to argue that large language models will essentially capture nonobvious latent legal epistemological biases in the training data. We believe that this conceptual schema based legal epistemological bias presents a novel

---

https://www.law.com/2023/07/25/a-welcome-boon-or-a-dreaded-replacement-paralegals-split-on-generative-ai/?utm_source=email&utm_medium=enl&utm_content=20230726&utm_campaign=morningminute&utm_term=law&oly_enc_id=7910E2562589G8R

[267] Lohr, Steve. "A.I. Is Coming for Lawyers, Again." The New York Times, April 10, 2023. https://www.nytimes.com/2023/04/10/technology/ai-is-coming-for-lawyers-again.html

[268] Keynes, John Maynard. *They Economic Possibilities of Our Grandchildren* (1930), noting that "We are being afflicted with a new disease of which some readers may not yet have heard the name, but of which they will hear a great deal in the years to come – namely, technological unemployment. This means unemployment due to our discovery of means of economising the use of labour outrunning the pace at which we can find new uses for labour."
See also: Jones, Steven E. *Against technology: From the Luddites to neo-Luddism*. Routledge, 2013.

[269] Samuel, Geoffrey. *Epistemology and method in law*. Routledge, 2016..

level of bias. Thus, under such a bias, certain conceptual schema will be preferred over others depending on the prompt and the underlying training data. This in turn will limit the decision-making capabilities of such models. Accordingly, while large language models are not explicitly designed to capture legal epistemological biases, they will nevertheless likely still exhibit biases that are rooted in schema of intelligibility-based patterns of legal reasoning observed in the text. These biases are in turn inherent in the way the legal texts reflect recognizable patterns of legal reasoning, which can be explained and explored using Berthelot's epistemological theory. In short, since legal text reflects legal reasoning patterns, and LLMs capture textual patterns, they will in turn capture and exhibit legal reasoning patterns and associated biases as reflected in the legal text.

3.2 Judicial Decision-making with ChatGPT: the Case of Judge Padilla

To illustrate the reality of a scenario where large language model-based AI is used to make legal decisions, it is worth starting with a motivating example. On January 30th, 2023, a Colombian judge, Juan Manuel Padilla, made history by being the first judge ever to use ChatGPT – or perhaps more accurately, the first judge to publicly acknowledge that he used ChatGPT – to help him write a ruling on a case.[270] The case concerned insurance coverage for the medical expenses of an autistic child whose family could not afford treatment and transportation costs.[271] Judge Padilla claims to have asked ChatGPT "to rule whether a health insurance company could deny paying fees for medical appointments, therapy and transportation for an autistic boy named Salvador, given his parents' limited income."[272] Apparently, Judge Padilla's ChatGPT prompt was: "Is an autistic minor exonerated from paying fees for their therapies?", among other questions. ChatGPT responded by stating the following: "Yes, this is correct. According to the regulations in Colombia, minors diagnosed with autism are exempt from paying fees for their therapies."[273]

---

[270] "Colombian Judge Says He Used Chatgpt in Ruling." The Guardian, February 3, 2023. https://www.theguardian.com/technology/2023/feb/03/colombia-judge-chatgpt-ruling
For other documented examples of the use of ChatGPT in court rulings see generally: Judges Using Generative AI - ChatGPT and Bing Chat Generative AI Legal Research Guide - LibGuides at University of Arizona Law Library. Accessed August 10, 2023. https://law-arizona.libguides.com/c.php?g=1301273&p=9671724

[271] Parikh, Purvish M., Dinesh M. Shah, and Kairav P. Parikh. "Judge Juan Manuel Padilla Garcia, ChatGPT, and a controversial medicolegal milestone." *Indian Journal of Medical Sciences* 75, no. 1 (2023): 3-8.

[272] Dailymail.com, Harriet Alexander. "Colombian Judge Uses Chatgpt to Make Decision in Legal First." Daily Mail Online, February 4, 2023. https://www.dailymail.co.uk/news/article-11712257/Colombian-judge-uses-ChatGPT-make-decision-legal-first.html

[273] *Ibid.*

Judge Padilla noted that "ChatGPT performs services previously provided by a secretary and did so "in an organized, simple and structured manner" which could "improve response times" in the justice system."[274] Furthermore he stated that he suspects that "many of [his] colleagues are going to join in this and begin to construct their rulings ethically with the help of artificial intelligence."[275] Of course, the ruling in question was written by Judge Padilla himself, who used Colombia's laws and precedents in his opinion. But what's striking is that the judicial decision-making process could be reduced to a simple question-and-answer procedure. In an interview to BluRadio,[276] Judge Padilla further noted that although "such programs could be useful to facilitate the drafting of texts" they should not be used "with the aim of replacing judges", adding that "by asking questions to the application, we do not stop being judges, thinking beings."[277]

With this example in mind, it becomes apparent that the notion of ChatGPT or a similar large language model-based AI being used for various legal tasks, including judicial decision-making, is not entirely unconceivable. Some scholars have argued that as long as the "robot judge" is "persuasive" then there should be no reason why such legal AI shouldn't be used for decision-making. [278] The question of legal AI-based decision-making is especially pertinent in light of the many costs involved in participating in the legal system, notwithstanding the pressures of case backlogs which paralyze many a judicial system worldwide.[279]

Indeed, if the current trends in AI development continue, it is not unforeseeable that in a couple of years Richard Susskind's predictions that due to AI "in the 2020s there will be many fundamental shifts in the legal sector", and that by 2036 "the legal profession will have changed beyond recognition," could indeed come to fruition.[280] But the important question raised by Judge Padilla's use of ChatGPT for legal decision-making

---

[274] *Ibid.*

[275] "Colombian Judge Uses Chatgpt in Ruling on Child's Medical Rights Case." CBS News, February 2, 2023. https://www.cbsnews.com/news/colombian-judge-uses-chatgpt-in-ruling-on-childs-medical-rights-case/

[276] León, Alejandro, and Blu Radio. "Sentencia La Tomé Yo, Chatgpt Respaldó Argumentación: Juez de Cartagena Usó Inteligencia Artificial." Blu Radio, February 2, 2023. https://www.bluradio.com/judicial/sentencia-la-tome-yo-chatgpt-respaldo-argumentacion-juez-de-cartagena-uso-inteligencia-artificial-pr30

[277] Abbott, Ryan, and Brinson S. Elliott. "Putting the Artificial Intelligence in Alternative Dispute Resolution: How AI Rules Will Become ADR Rules." *Amicus Curiae* 4, no. 3 (2023): 685-706. See also: Dailymail.com, Harriet Alexander. "Colombian Judge Uses Chatgpt to Make Decision in Legal First." Daily Mail Online, February 4, 2023. https://www.dailymail.co.uk/news/article-11712257/Colombian-judge-uses-ChatGPT-make-decision-legal-first.html

[278] See generally: Volokh, Eugene. "Chief justice robots." *Duke LJ* 68 (2018): 1135.

[279] Susskind, Richard. "Online courts and the future of justice." (2019).

[280] Susskind, Richard. *Tomorrow's lawyers: An introduction to your future*. Oxford University Press, 2013.

remains unanswered: could we potentially have actual cases decided with the help of AI?[281] And if so, what are the issues in large language models that might reflect on such judicial decision-making?

3.2.1 An Epistemological Question

Although the reality of legal decision-making being done by AI is likely beyond the current state of the art for large language model technologies,[282] it should be noted that from the perspective of legal epistemology the question itself is neither a speculative nor a trivial one. Indeed, as some legal scholars have pointed out, the question of artificial intelligence-based judging is of central importance to understanding both legal knowledge and legal epistemology as such. For instance, in a recent article Professor Samuel pointed out that "the moment one starts to talk about computer or robot judges one might also begin to appreciate the extent to which legal theory and legal education have in truth failed to expose the tensions [that have plagued legal knowledge] in a way that actually impacts on legal learning and traditional (doctrinal) legal scholarship."[283] Furthermore, in "Epistemology and Method in Law", Samuel also notes that "with the advent of artificial intelligence (AI), the question of what it is to have knowledge of law has become important inasmuch as the existing theories of legal knowledge have proved inadequate."[284] Further in the past, in a 1991 article on a similar issue, Professor Samuel noted that "one positive result that could well emerge from all the research into such [Artificial Intelligence] modelling is a fundamental reassessment of legal theory."[285]

---

[281] Susskind, Richard. *Expert systems in law*. Oxford University Press, Inc., 1987.
See: Sourdin, Tania. "Judge v Robot?: Artificial intelligence and judicial decision-making." *University of New South Wales Law Journal, The* 41, no. 4 (2018): 1114-1133.
See also: Wang, Nu. "'"Black Box Justice": Robot Judges and AI-based Judgment Processes in China's Court System." In *2020 IEEE International Symposium on Technology and Society (ISTAS)*, pp. 58-65. IEEE, 2020.
See also: Xu, Zichun. "Human Judges in the era of artificial intelligence: challenges and opportunities." *Applied Artificial Intelligence* 36, no. 1 (2022): 2013652.

[282] Primarily because of the issue of hallucination of unintended text which can have detrimental consequences
See: Ji, Ziwei, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. "Survey of hallucination in natural language generation." *ACM Computing Surveys* 55, no. 12 (2023): 1-38.
The issue of large language model hallucinations will also be discussed below
See also: Bohannon, Molly. "Lawyer Used Chatgpt in Court-and Cited Fake Cases. A Judge Is Considering Sanctions." Forbes, June 9, 2023.
https://www.forbes.com/sites/mollybohannon/2023/06/08/lawyer-used-chatgpt-in-court-and-cited-fake-cases-a-judge-is-considering-sanctions/?sh=33cdc6627c7f

[283] Samuel, Geoffrey. "Is Legal Knowledge Regressing (Thanks to AI)?." *Amicus Curiae* 4, no. 3 (2023): 719-750.

[284] Samuel, Geoffrey. *Epistemology and method in law*. Routledge, 2016.

[285] Samuel, Geoffrey. "The challenge of artificial intelligence: can Roman law help us discover whether law is a system of rules?." *Legal Studies* 11, no. 1 (1991): 24-46.

Therefore, the question of AI judging and legal decision-making is actually another way of asking a legal epistemological question of what it means "to judge" or "to have legal knowledge" or to "know the law."[286] Accordingly, if AI was somehow capable of representing legal knowledge and thus capable of making legal decisions, what exactly would that legal knowledge be?[287] And are the current large language model-based approaches sufficient for this task? We can therefore now see how a seemingly hypothetical question on the nature of AI, large language models and their application in the legal domain turns into something much more complex and fundamental to legal theory. It is therefore unsurprising that, according to Susskind, this seemingly speculative question of "robot judges" actually "reaches into the very core of jurisprudence and philosophy."[288]

3.3 In Search of Legal Reasoning

Before discussing the relevant limitations of large language models, we must first examine what legal reasoning actually entails and whether it is possible for legal reasoning to be represented by these AI systems.

A fundamental problem of legal reasoning – apart from the debates involved in defining what legal reasoning actually is[289] – is the reality that more often than not, legal cases can be decided either way.[290] Furthermore, there are no objective criteria in law for determining the validity of a certain interpretation, or as the jurist Hans Kelsen states "all methods of interpretation developed so far lead only to a possible, not a necessary, result, never to one which is alone correct."[291] For example, simply looking at the cases at the Supreme Court, one can see that judges, despite dealing with identical facts and sets of laws, oftentimes tend to disagree, arguably along political lines.[292] Although not all cases are indeed "hard cases," these types of cases tend to illustrate most clearly the problem of why legal reasoning is fundamentally problematic.

Gottfried Wilhelm Leibniz was also concerned with the problems of legal reasoning throughout his life. Before he became famous for his mathematical discoveries, Leibniz

---

[286] Samuel, Geoffrey. *Epistemology and method in law*. Routledge, 2016.
[287] *Ibid.*
[288] Susskind, Richard. *Expert systems in law*. Oxford University Press, Inc., 1987., p. 44.
[289] See: Alexander, Larry, and Emily Sherwin. "Demystifying legal reasoning." (2008).
See also: Samuel, Geoffrey. "Can legal reasoning be demystified?." *Legal Studies* 29, no. 2 (2009): 181-210.
[290] Kelsen, Hans. *Pure theory of law*. Univ of California Press, 1967.
Dworkin, Ronald. "Hard cases." *Harv. L. Rev.* 88 (1974): 1057.
[291] Kelsen, Hans. *Pure theory of law*. Univ of California Press, 1967. p 352
[292] Martin, Andrew D., and Kevin M. Quinn. "Dynamic ideal point estimation via Markov chain Monte Carlo for the US Supreme Court, 1953–1999." *Political analysis* 10, no. 2 (2002): 134-153.

completed his dissertation for a Doctorate in Law which was titled "*Disputatio Inauguralis de Casibus Perplexis in Jure*" or "Inaugural Disputation on Ambiguous Legal Cases" in 1666, at the age of twenty.[293] In his 1679 essay titled "Towards a numerical universal language", which, judging by the title alone, anticipated much of what we would call natural language processing today, Leibniz states the following observation:

> "Two disputing persons seem to me to resemble two merchants who have long been in each other's debt on many counts, but who never wanted to perform an examination [of their accounts] by means of a general balance. Instead, each keeps exaggerating in a different way his respective credits and the truth and magnitude of his respective commitments (nominum) (i.e., of their debts): obviously they will never put an end to their contest. We should not wonder that this has happened so far in many controversies where the issue is not transparent (that is, subject to numbers)."[294]

The last sentence offers a hint as to Leibniz's view on how a judge is supposed to reason and come to a decision on these types of cases, namely, by using numbers and deductive reasoning as is the case in mathematics.[295] Earlier, in 1669, in "*Elementa juris naturalis*" or "The Elements of Natural Law"[296], Leibniz similarly stated the following:

> "The doctrine of Right belongs to those sciences which depend on definitions and not on experience and on demonstrations of reason and not of sense; they are problems of law, so to speak, and not of fact. For since justice consists in a kind of congruity and proportionality, we can understand that something is just even if there is no one who practices it or upon whom it is practiced. Just so the relations of numbers are true even if there were no one to count and nothing to be counted, and we can predict that a house will be beautiful, a machine efficient, or a commonwealth happy, if it comes into being, even if it should never do so. We need not wonder, therefore, that the principles of these sciences possess eternal truth."[297]

---

[293] Gordley, James. The Jurists: A Critical History. United Kingdom: OUP Oxford, 2013.
See also: Arthur, Richard TW. *Leibniz*. John Wiley & Sons, 2014.
[294] Dascal, Marcelo. *Gottfried Wilhelm Leibniz: the art of controversies*. Edited by Quintín Racionero, and Adelino Cardoso. Springer Science+ Business Media BV, 2006.
[295] Gordley, James. *The jurists: a critical history*. OUP Oxford, 2013.
[296] Armgardt, Matthias. "Leibniz as legal scholar." *Fundamina: A Journal of Legal History* 2014, no. si-1 (2014): 27-38.
See also: Hoeflich, Michael H. "Law & geometry: Legal science from Leibniz to Langdell." *American Journal of Legal History* 30, no. 2 (1986): 95-121.
[297] Leibniz, Gottfried Wilhelm. "Elements of Natural Law." In *Philosophical papers and letters*, pp. 131-138. Springer Netherlands, 1989. 2nd edition Translated by: Leroy E. Loemker
See also: Gordley, James. *The jurists: a critical history*. OUP Oxford, 2013. For another translation of this quote

It is not mere numbers however that Leibniz was interested in, but how to represent words and legal concepts. He continues, stating that:

> "Since the doctrine of Law is a science, and the basis of science is demonstration, and definition is the principle of demonstration, it follows that we must first of all investigate the definitions of the words *Right*, *just*, and *justice*, that is, the clear ideas by which we usually estimate the truth of propositions or of the right use of words in speech, even when we do not know we are doing so. The method of our investigation is to gather the more important and distinctive examples of the use of these terms and to set up some meaning consistent with these and other examples."[298]

Discussions on Leibniz in the context of law and AI are of fundamental importance because they reveal arguably the first historically recorded attempts at "computerizing law."[299] Leibniz's solution to lack of objectivity in legal reasoning was to develop a methodology for legal decision-making based on mathematics and geometry, known as the *mos geometricus*[300] or *mos mathematicus*,[301] where *mos* is translated from Latin as "way" or "manner." As Deakin and Markou point out: "Leibniz believed that it was possible to develop a consistent system of logic, language and mathematics using an alphabet of unambiguous symbols that could be manipulated according to mechanical rules."[302] Furthermore, Gordley states that: "like Suárez, [Leibniz and Wolff] sought immutable principles. Like Descartes, they wished to establish them by deductive logic. They believed that they had deduced immutable principles of law from concepts invariably attached to human nature."[303] The impact and influence of the *mos geometricus* was enormous as it served as the methodological and a theoretical foundation for the codification projects around the world that subsequently followed.[304] In the words of the German jurist von Savigny, who was writing in 1814, almost two centuries after Leibniz: "In every triangle, namely, there are certain data, from the relations of which all the rest are necessarily deducible: thus, given two sides and the included angle, the whole triangle

---

[298] *Ibid.*

[299] Samuel, Geoffrey. "Is Legal Knowledge Regressing (Thanks to AI)?." *Series 2 Vol. 4 Amicus Curiae* (2023): 719, where Professor Samuel however points out that "Leibniz had Roman law in mind", so the question is even older than Leibniz.
See also: Deakin, Simon, and Christopher Markou, eds. *Is law computable?: critical perspectives on law and artificial intelligence*. Bloomsbury Publishing, 2020.

[300] Gordley, James. *The jurists: a critical history*. OUP Oxford, 2013.

[301] Samuel, Geoffrey. "Is Legal Knowledge Regressing (Thanks to AI)?." *Series 2 Vol. 4 Amicus Curiae* (2023): 719.

[302] Deakin, Simon, and Christopher Markou, eds. *Is law computable?: critical perspectives on law and artificial intelligence*. Bloomsbury Publishing, 2020.

[303] Gordley, James. *The jurists: a critical history*. OUP Oxford, 2013.
See also: Gordley's discussions on Christian Wolff, who was another philosopher, mathematician and jurist working to develop the *mos geometricus*

[304] For a rich discussion of this see: Gordley, James. *The jurists: a critical history*. OUP Oxford, 2013.

is given. In like manner, every part of our law has points by which the rest may be given: these may be termed the leading axioms."[305] What these discussions demonstrate is that the problem of computerization and systematization of law is indeed an ancient one. But what's more important is that, in fact, the assumptions underlying the current development of legal AI are indeed fundamentally similar to the deductive assumptions of Leibniz and the *mos geometricus* project. Indeed, Deakin and Markou argue, the current approaches still "rest on the Leibnizian-Langdellian assumption that there is a purified essence to law and legal reasoning there to be mathematised."[306] Furthermore, as Samuel notes "the very issue of AI and law has, then, the effect of exposing all the contradictions and tensions that have "plagued" legal knowledge not just in recent times but in past centuries as well."[307]

But if legal reasoning is not some deductive process about rules and cases, then what is it?[308] The problem is that, unfortunately, legal scholarship itself doesn't really have a singular and satisfactory answer to this question. For instance, as Samuel points out "the pessimism [underpinning the question of whether law is computable] springs primarily from the woeful state of epistemological thinking in law."[309] Thus, the answer to the question we posed earlier on whether it is possible for legal reasoning to be represented by these large language models is to ask another question, namely, what do we even mean by legal reasoning?

3.3.1 Berthelot's Schemas of Intelligibility[310]

Luckily, there is a theory which can help us explore how decision-making actually takes place in the legal domain. A perspective that we adopt in the present work is Professor Samuel's extension of Berthelot's "schemes of intelligibility"[311] to the legal domain.[312] Samuel points out that "as a result of work published by the French social science epistemologist Jean-Michel Berthelot, there is now a real possibility of locating

[305] von Savigny, Friedrich Carl. *Of the Vocation of Our Age for Legislation and Jurisprudence*. Littlewood and Company, 1831. Translated by: Abraham Hayward, p 38-39
See also: Gordley, James. *The jurists: a critical history*. OUP Oxford, 2013.
[306] Deakin, Simon, and Christopher Markou, eds. *Is law computable?: critical perspectives on law and artificial intelligence*. Bloomsbury Publishing, 2020.
[307] Samuel, Geoffrey. "Is Legal Knowledge Regressing (Thanks to AI)?." *Series 2 Vol. 4 Amicus Curiae* (2023): 719.
[308] Samuel, Geoffrey. *Epistemology and method in law*. Routledge, 2016.
[309] Samuel, Geoffrey. "Is Legal Knowledge Regressing (Thanks to AI)?." *Series 2 Vol. 4 Amicus Curiae* (2023): 719..
[310] Berthelot, Jean-Michel. *Épistémologie des sciences sociales*. PUF, 2001.
Also referred to as schemes
For a detailed explanation of the schemes of intelligibility as applied in the legal domain see generally: Samuel, Geoffrey. *Epistemology and method in law*. Routledge, 2016.
[311] Berthelot, Jean-Michel. *Épistémologie des sciences sociales*. PUF, 2001.
[312] Chatper 8, Samuel, Geoffrey. *Epistemology and method in law*. Routledge, 2016.

legal reasoning within a wider epistemological framework."[313] Before moving onward, it is worth pointing out that Berthelot's work is published in French and, unfortunately, not translated into English, so I rely here on Professor Samuel's extensive and invaluable research, translations and applications of Berthelot's work.

The value of Berthelot's schemes of intelligibility lies in their explanatory power, particularly when it comes to legal reasoning in hard cases, and indeed, according to Samuel, "one can often discern [the relevance of the schemes of intelligibility to legal reasoning] when there are differences between judges which may occur in the same court—dissenting opinions—or between two courts when, say, the Supreme Court judges overrule a decision of the Court of Appeal."[314]

In total there are six basic schemes of intelligibility proposed by Berthelot, in addition to their interactions:[315] the causal scheme, the functional scheme, the structural scheme, the hermeneutical scheme, the actional scheme, and the dialectical scheme.[316] According to Samuel: "the employment of different schemes, or mixture of schemes, results in different knowledge."[317] In the table below, Samuel's adoption and translation of Berthelot's schemes of intelligibility, is provided, together with some of his important comments and applications to law.

---

[313] *Ibid.*

[314] Samuel, Geoffrey. "Is Legal Knowledge Regressing (Thanks to AI)?." *Series 2 Vol. 4 Amicus Curiae* (2023): 719.

[315] The nature of schemes of intelligibility can be analogized to Vladimir Propp's narrative structural analysis of Russian folk tales, whereby he was able to deduce thirty-one basic structural elements, or "functions" in a fairy tale, for instance "a hero receives a magic tool", etc.
See generally: Propp, Vladimir. *Theory and history of folklore*. Vol. 5. U of Minnesota Press, 1985.
See also: Propp, Vladimir. *Morphology of the Folktale*. University of Texas Press, 1968.

[316] For an in-depth discussion on Berthelot's schemes of intelligibility as applied to law see:
Chatper 8, Samuel, Geoffrey. *Epistemology and method in law*. Routledge, 2016.
See also: Samuel, Geoffrey. *Rethinking legal reasoning*. Edward Elgar Publishing, 2018.
See also: Samuel, Geoffrey. "Can legal reasoning be demystified?." *Legal Studies* 29, no. 2 (2009): 181-210.
See also: Samuel, Geoffrey. "Is law really a social science? A view from comparative law." *The Cambridge Law Journal* 67, no. 2 (2008): 288-321.

[317] Samuel, Geoffrey. "Is Legal Knowledge Regressing (Thanks to AI)?." *Series 2 Vol. 4 Amicus Curiae* (2023): 719.

Schemes of Intelligibility	Applications to Legal Reasoning
I. *Causal* scheme (if *x*, then *y* or *y = f(x)*);[318]	1. "The phenomenon of damage (A) must be dependent on the act of the defendant (B)"[319] 2. "Tort lawyers spend much time applying causal and actional methods in the analysis of case law problems."[320]
II. *Functional* scheme (S → X → S, where one phenomenon X is analysed from the position of its function – X → S – in a given system);[321]	1. "Legal rules and concepts are understood in terms of their function"[322] and "What is the purpose of the rule in question?"[323] 2. Functional methods include "policy reasoning or the application of the mischief rule in statutory interpretation."[324]

[318] Samuel, Geoffrey. "Is law really a social science? A view from comparative law." *The Cambridge Law Journal* 67, no. 2 (2008): p 298
[319] Samuel, Geoffrey. *Epistemology and method in law*. Routledge, 2016. p 303
[320] Samuel, Geoffrey. "Is law really a social science? A view from comparative law." *The Cambridge Law Journal* 67, no. 2 (2008)
[321] *Ibid.*
[322] Samuel, Geoffrey. *Epistemology and method in law*. Routledge, 2016. p 306
[323] Samuel, Geoffrey. "Can legal reasoning be demystified?." *Legal Studies* 29, no. 2 (2009): 181-210.
[324] Samuel, Geoffrey. "Is law really a social science? A view from comparative law." *The Cambridge Law Journal* 67, no. 2 (2008): p 298

III. *Structural* scheme (where X results from a system founded, like language, on disjunctive rules, A or not A);[325]	1. "Just as the 'red light' and 'green light' can be seen as part of a code where one draws its meaning only from its relation to the other…so the legal subject has meaning only in relation to the legal object and vice versa."[326] 2. "The [structural] scheme can be said to underlie the use of analogy and metaphor in legal reasoning… For example, ships have little in common with cars and the sea has little in common with roads; nevertheless, an analogy has been drawn in law between damage done by ships to an adjacent beach and damage done by cars to property adjacent to a highway…What is similar is the symmetry of the structure, that is to say the isomorphic quality of the two situations."[327]

[325] *Ibid.*
[326] Samuel, Geoffrey. *Epistemology and method in law*. Routledge, 2016. p 307
[327] *Ibid.* p 307

IV. *Hermeneutical* scheme (where X is the symptom, the expression of an underlying signification to be discovered through interpretation);[328]	1. "Fact A is regarded as a signifier of something deeper, the signified (fact B)"[329] For instance: "Evidence (fact A, the signifier) discloses a hostility on behalf of the claimant (fact B, the signified) that goes beyond what is reasonable… and thus signifies something deeper."[330] 2. "In the social sciences, it is the human facts which act as the signifier, and it is for the particular science to explain the meaning of these signs. Facts, in other words, are only a surface phenomenon; what they need is a science to explain their meaning, their significance."[331] 3. "Statutory interpretation in the common law and even the analysis of precedents might equally be labelled ars hermeneutica."[332] Thus, hermeneutical analysis asks "What did the legislator mean or what was the will of the legislator or testator?"[333] Furthermore, "the causal, functional, structural, actional and dialectical schemes can all be viewed as signs which on deeper analysis reveal hidden meanings; they can, in other words, be read hermeneutically."[334]

[328] Samuel, Geoffrey. "Is law really a social science? A view from comparative law." *The Cambridge Law Journal* 67, no. 2 (2008): p 298
[329] Samuel, Geoffrey. "Can legal reasoning be demystified?." *Legal Studies* 29, no. 2 (2009): 181-210
[330] *Ibid.*
[331] Samuel, Geoffrey. *Epistemology and method in law*. Routledge, 2016. p 309
[332] *Ibid.* p 311
[333] Samuel, Geoffrey. "Is law really a social science? A view from comparative law." *The Cambridge Law Journal* 67, no. 2 (2008): p 298
[334] Samuel, Geoffrey. *Epistemology and method in law*. Routledge, 2016. p 312

V. *Actional* scheme (where X is the outcome, within a given space, of intentional actions);[335]	1. "The emphasis of the actional scheme is on the actor…whose behaviour is to be understood in terms of intention and effect…the 'reasonable man', the 'contracting party' and so on are some of the better-known examples. Indeed, the whole of Western criminal law could be said to be premised on this scheme."[336] 2. In an actional analysis "facts are viewed in terms of constructed individual actors and their intentions"[337] thus, "a phenomenon is to be rationalised through the three elements of intention, act and effect."[338]
VI. *Dialectical* scheme (where X is the necessary outcome of the development of internal contradictions within a system)[339]	1. "Dialectical reasoning can be found in conceptual oppositions such as the contradiction between the rights, or interests, of the plaintiff on the one hand and the defendant on the other… A notion like the owner's rights thus presupposes the coexistence of a contrary notion such as the tenant's rights."[340] 2. For example, a dispute can be reduced to an "interest of the public at large [fact A]; and the interest of a private individual [fact non-A]"[341] 3. "The main importance of dialectics as a scheme of intelligibility is to be found in its formal organising quality. It may not explain the internal movement or evolution towards a solution, but it structures oppositions in such a way as to channel a reasoning or argumentation process towards, if not a compromise, then a gamelike conclusion whose tactics and strategies are constructed through the employment of other schemes of intelligibility."[342]

[335] Samuel, Geoffrey. "Is law really a social science? A view from comparative law." *The Cambridge Law Journal* 67, no. 2 (2008): p 298
[336] *Ibid.*
[337] Samuel, Geoffrey. "Can legal reasoning be demystified?." *Legal Studies* 29, no. 2 (2009): 181-210
[338] Samuel, Geoffrey. *Epistemology and method in law*. Routledge, 2016. p 313
[339] Samuel, Geoffrey. "Is law really a social science? A view from comparative law." *The Cambridge Law Journal* 67, no. 2 (2008): p 298
[340] *Ibid* p 316.
[341] Samuel, Geoffrey. "Can legal reasoning be demystified?." *Legal Studies* 29, no. 2 (2009): 181-210.
[342] Samuel, Geoffrey. *Epistemology and method in law*. Routledge, 2016. p 317

These schemes of intelligibility can also interact with each other, and the adoption of different combinations of these conceptual schemas thus results in different legal consequences.[343] Judges thus can easily switch between different schemes or combine them to reach a desired conclusion.[344] This explains the puzzle of how judges so often disagree despite dealing with identical sets of laws and facts.[345] Thus Berthelot's schemes of intelligibility offer a novel lens to understanding legal reasoning by allowing us to examine how different legally valid decisions can be constructed using the different schemes.

Fundamentally, it can thus be said that legal reasoning and legal decision-making is essentially perspectivist in nature,[346] in the sense that the laws and facts can be seen as mere tools for predetermined conclusions that a judge has in their mind. We can call this feature of legal decision-making "retrocausality" in the sense that the effect, namely the legal decision, creates its own causes, or the arguments for said decision, thus, paradoxically, the effect precedes its own causes. [347] In other words, a judge who has the desired conclusion in mind before the fact can utilize the various schemes of intelligibility, combining, emphasizing, discounting certain facts, laws and situations, etc. The complexity of this process is likely why a deductive Leibnizian approach to legal reasoning has not proven to be successful.

There is confirmation for this perspectivist view of decision-making from other disciplines as well. For instance, Haidt's thesis underpinning his moral foundations theory is that "moral reasoning was mostly just a post hoc search for reasons to justify the

---

[343] Samuel, Geoffrey. *Epistemology and method in law*. Routledge, 2016, see Chatpter 8 generally

[344] This is not too dissimilar to Schopenhauer's discussion on conceptual spheres in the context of the art of persuasion.
See: Schopenhauer, Arthur*: 'The World as Will and Representation': Volume 1* (1818). p 72-74. Cambridge University Press, 2010 ed. Trans: Judith Norman, Alistair Welchman and Christopher Janaway

[345] For a Nietzschean philosophical perspective on non-existence of non-perspectival objectivity, see: Hales, Steven D., and Rex Welshon. *Nietzsche's perspectivism*. University of Illinois Press, 2000.
See generally: Nietzsche, Friedrich. *The will to power*. Vintage, 1968.

[346] Nietzsche, Freidrich, *Will to Power* (1901/1968) translated by Kaufman, §481, stating that: "Against positivism, which halts at phenomena -"There are only *facts" - I* would say: No, facts is precisely what there is not, only interpretations. We cannot establish any fact "in itself": perhaps it is folly to want to do such a thing. "Everything is subjective," you say; but even this is interpretation. The "subject" is not something given, it is something added and invented and projected behind what there is. Finally, is it necessary to posit an interpreter behind the interpretation? Even this is invention, hypothesis. interpretations."

[347] Zizek, Slavoj. *The sublime object of ideology*. Verso Books, 2009
See also: Žižek, Slavoj. *Event: Philosophy in transit*. Penguin UK, 2014.
See also: Dupuy, Jean-Pierre *Petite métaphysique des tsunamis*, Paris: Editions du Seuil 2005, p. 19.

judgments people had already made."[348] Haidt's view itself is based on Hume's famous argument that "reason is and ought only to be the slave of the passions."[349] Schopenhauer's discussions on the art of persuasion are also of value here as well, as he claims that any persuasive argument can be made by building up arguments via the relationship between what he calls "conceptual spheres", and importantly, these arguments are constructed "according to [the speaker's] original intention."[350] Furthermore, Montaigne also stated that:

> "Men of intelligence notice more things and view them more carefully, but they [interpret] them; and to establish and substantiate their interpretation, they cannot refrain from altering the facts a little. They never present things just as they are but twist and disguise them to conform to the point of view from which they have seen them; and to gain credence for their opinion and make it attractive, they do not mind adding something of their own, or extending and amplifying."[351]

Berthelot's schemes of intelligibility allow us to detect in a systematic fashion exactly how is it that these so-called men of intelligence are able substantiate their points of view. Putting aside the broader philosophical implications of Berthelot's theory for law, it is worth briefly discussing the actual applied nature of this theory. For instance, Professor Samuel applies the schemes of intelligibility as follows, noting that:

> "A legislative rule can be understood in a structural way, that is to say in relation to the other rules in the same piece of legislation – or in respect of a wider conceptual structure – or in a functional way. What is the purpose of the rule in question? In order to understand this purpose or function the interpreter may adopt a hermeneutical approach, asking what does the language signify in respect to the will of the legislator. Alternatively, the interpreter might apply a dialectical scheme by setting up a contradicting alternative."[352]

To provide a further example of how Berthelot's schemes of intelligibility are applied in concrete legal cases, let us consider a famous and controversial English tort law case of *Gregg v Scott* which concerned the doctrine of the loss of chance.[353] In this

---

[348] Haidt, Jonathan. *The righteous mind: Why good people are divided by politics and religion*. Vintage, 2012.
[349] Hume, David. *A treatise of human nature*. Clarendon Press, 1896.
[350] Schopenhauer, Arthur. *The world as will and representation, vol. 1*. Vol. 1. Courier Corporation, 2012., p 72
[351] de Montaigne, Michel (1595). *"Of Cannibals". Essays.* Trans: Cohen, John M. Penguin (195y)
[352] Samuel, Geoffrey. "Can legal reasoning be demystified?." *Legal Studies* 29, no. 2 (2009): 181-210.
[353] *Gregg v Scott* [2005] UKHL 2
For a discussion on the application of Berthelot's schemes of intelligibility to concrete legal cases see the discussion on the case of *Miller v Jackson* in Samuel, Geoffrey. "Can legal reasoning be demystified?." *Legal Studies* 29, no. 2 (2009): 181-210.

case, which was decided at the UK Supreme Court in 2005, Mr. Gregg's cancer was negligently misdiagnosed as a benign lump, and thus Mr. Gregg did not receive the necessary treatment on time which significantly reduced his chance of survival. Consider the application of the causal scheme to these facts: did the doctor, by his misdiagnosis, cause "actual damage in the form of the enlarged tumor?"[354] Or did the doctor, by his misdiagnosis, merely cause a reduction of chance of survival from 42% to 25%, which, at 42%, on the balance of probabilities means that Mr. Gregg would likely have not recovered from the cancer anyway?[355] If we further add a functional analysis, we can also ask what the purpose or function of the doctrine of causation in UK medical negligence law is. Is it perhaps to protect the financial state of the National Health Service and various medical insurance companies by limiting the number of possible claims?[356] Furthermore, a dialectical scheme can reduce the problem to helping a person who suffered from a serious case of medical negligence versus the financial interests of the taxpayer-funded National Health Service.[357] An actional analysis might examine the role of the doctor and his duty to the patient and whether that includes "preventing a reduction in the appellant's prospects of a successful [medical] recovery."[358] A hermeneutic analysis might emphasize that, despite the misdiagnosis and the alleged harms suffered, Mr. Gregg is still alive eight years after the fact, and thus, his claim of lost years may be unjustified.[359]

Importantly, all of these applications of schemes of intelligibility were taken *directly from the text* of the decision, and thus, it is clear that scheme of intelligibility-based patterns of legal reasoning are contained and observable within the text. Thus, if these schemes of intelligibility exist in the text, can large language models capture these reasoning patterns, and if so, what are the epistemological problems that may arise as a result? To restate the question in a slightly different way, what are the legal epistemological problems that one might face when using AI for legal decision-making, particularly in the context of large language models' propensity towards capturing bias in the training data and their uncertain reasoning capabilities? [360]

---

[354] Fordham, Margaret. "Loss of Chance-A Lost Opportunity." *Sing J. Legal Stud.* (2005): 204. *Gregg v Scott* [2005] UKHL 2, para 95
[355] Fordham, Margaret. "Loss of Chance-A Lost Opportunity." *Sing J. Legal Stud.* (2005): 204.
[356] *Gregg v Scott* [2005] UKHL 2, para 90
[357] *Ibid.* para 43
[358] *Ibid*. para 102
[359] *Ibid.* para 133
[360] I shall not discuss the social, legal and political objections as much has been written on this subject. Rather, my focus is on legal epistemology and the discussion of the limited pattern detection-based abilities of large language models in capturing legal knowledge in their current state particularly in the context of the underlying biases in the data. For discussions on social, legal and political consequences of AI judging see for instance: Morison, John, and Adam Harkens. "Re-engineering justice? Robot judges, computerised courts and (semi) automated legal decision-making." *Legal Studies* 39, no. 4 (2019): 618-635., etc.
See also: Sourdin, Tania. "Judge v Robot?: Artificial intelligence and judicial decision-making." *University of New South Wales Law Journal, The* 41, no. 4 (2018): 1114-1133.

3.4 Large Language Models and the Issue of Bias

Since first appearing in 2018,[361] large language models – which includes models like GPT-3.5 and GPT-4 upon which the chatbot ChatGPT is built – have arguably revolutionized NLP.[362] Therefore, it is worth briefly examining what is meant by the term "large language model."

Firstly, the "language model" part of "large language model" refers to a natural language processing task that tries to "model the generative likelihood of word sequences, so as to predict the probabilities of future (or missing) tokens"[363] – which is actually not too dissimilar to gap filling exercises kindergarten students or language learners do to learn a language.[364] Just like a kindergarten student who tries to successfully complete the fill in the gap exercise, the language modeling task relies on taking account of the contextual information in a sentence. This is based on the "distributional hypothesis" which is a theory that "words that occur in similar contexts tend to have similar meanings."[365] This theory can be restated in the infamous words of

---

Hilliard, Airlie, Nigel Guenole, and Franziska Leutner. "Robots are judging me: Perceived fairness of algorithmic recruitment tools." *Frontiers in Psychology* 13 (2022): 940456.

[361] Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. "Bert: Pre-training of deep bidirectional transformers for language understanding." *arXiv preprint arXiv:1810.04805* (2018).

[362] Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. "Attention is all you need." *Advances in neural information processing systems* 30 (2017).
See also: Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. "Bert: Pre-training of deep bidirectional transformers for language understanding." *arXiv preprint arXiv:1810.04805* (2018).
See also: Radford, Alec, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. "Improving language understanding by generative pre-training." (2018).
See also: Brown, Tom, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D. Kaplan, Prafulla Dhariwal, Arvind Neelakantan et al. "Language models are few-shot learners." *Advances in neural information processing systems* 33 (2020): 1877-1901.
For a survey of the history of Large Language Models, see generally: Zhao, Wayne Xin, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min et al. "A survey of large language models." *arXiv preprint arXiv:2303.18223* (2023).

[363] Zhao, Wayne Xin, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min et al. "A survey of large language models." *arXiv preprint arXiv:2303.18223* (2023).
See also: Jurafsky, Daniel, and James H. Martin. "Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition." 2023 ed.

[364] See: Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. "Distributed representations of words and phrases and their compositionality." *Advances in neural information processing systems* 26 (2013).
See also: Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. "Efficient estimation of word representations in vector space." *arXiv preprint arXiv:1301.3781* (2013).

[365] Jurafsky, Daniel, and James H. Martin. "Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition." 2023 ed.

linguist J.R. Firth who said that "you shall know a word by the company it keeps.'"[366] Thus, conceptually, language modeling works in a similar way to a human who, upon reading an unknown word in a sequence of text, tries to gain an understanding of that word by examining the context words surrounding that word.[367]

On the other hand, the "large" part of the "large language models" refers to the fact that these models tend to contain billions, and even trillions of parameters. For instance, the first large language model, Bidirectional Encoder Representations from Transformers (BERT) has 340 million parameters,[368] the GPT-3 model has 175 billion parameters,[369] and the recently released PanGu-Σ for Chinese language contains 1.085 trillion parameters.[370] Parameters themselves can be thought of as a model's "internal adjustable "knobs" that reduce the error measure, (or distance) between the output scores and the desired pattern of scores."[371] In essence, parameters are aspects of the model that need to be updated and optimized in order to give the lowest error to accomplish a task at hand, and in the case of language modeling, that entails finding the best parameters for predicting the next word in a sequence.

Of course, the sheer number of parameters involved fundamentally complicates the interpretation of the meaning of these parameters and their interactions, and what exactly the model learns from the text,[372] which is why these large language models are often referred to as being "black-box architectures."[373] Thus, a fundamental issue with large language models is that although we are able to observe the output of a chatbot like ChatGPT, nevertheless, it's incredibly hard to tell how exactly this output is actually

---

[366] Firth, John. "A synopsis of linguistic theory, 1930-1955." *Studies in linguistic analysis* (1957): 10-32.
[367] Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. "Efficient estimation of word representations in vector space." *arXiv preprint arXiv:1301.3781* (2013).
[368] Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. "Bert: Pre-training of deep bidirectional transformers for language understanding." *arXiv preprint arXiv:1810.04805* (2018).
[369] Brown, Tom, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D. Kaplan, Prafulla Dhariwal, Arvind Neelakantan et al. "Language models are few-shot learners." *Advances in neural information processing systems* 33 (2020): 1877-1901.
[370] Ren, Xiaozhe, Pingyi Zhou, Xinfan Meng, Xinjing Huang, Yadao Wang, Weichao Wang, Pengfei Li et al. "PanGu-{\Sigma}: Towards Trillion Parameter Language Model with Sparse Heterogeneous Computing." *arXiv preprint arXiv:2303.10845* (2023).
See also: Fedus, William, Barret Zoph, and Noam Shazeer. "Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity." *The Journal of Machine Learning Research* 23, no. 1 (2022): 5232-5270.
[371] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." *nature* 521, no. 7553 (2015): 436-444.
[372] Clark, Kevin, Urvashi Khandelwal, Omer Levy, and Christopher D. Manning. "What does bert look at? an analysis of bert's attention." *arXiv preprint arXiv:1906.04341* (2019).
[373] Peng, Baolin, Michel Galley, Pengcheng He, Hao Cheng, Yujia Xie, Yu Hu, Qiuyuan Huang et al. "Check your facts and try again: Improving large language models with external knowledge and automated feedback." *arXiv preprint arXiv:2302.12813* (2023).

being generated and what is really going on with the parameters behind the scenes and under the hood. This problem was likely anticipated by Turing, who claimed that "an important feature of a learning machine is that its teacher will often be very largely ignorant of quite what is going on inside."[374] This ignorance of what is going on inside has in turn led to the development of an entire new subdiscipline in natural language processing dedicated precisely to studying the parameters in these large language models, which is known as BERTology.[375] Adding to this fundamental lack of interpretability of the parameters, the incredibly large financial and computational costs involved in training these large language models[376] and the storage costs involved in storing the training data, and even the environmental costs due to electricity usage and carbon footprint,[377] have created a situation where only large tech companies such as Google, OpenAI and Meta are even able to develop and release these large language models, which has led Luitse and Denkena to pertinently argue that "the growing size of language models accelerates monopolization in the digital economy."[378]

A further complication is the question of bias of large language models.[379] Since large language models use large amounts of online text data as their training data, they tend to capture relationships which tend to be negative, harmful, toxic and biased.[380] For instance, large language models seem to learn offensive racial,[381] gender,[382]

[374] Turing, Alan M. *Computing machinery and intelligence*. Springer Netherlands, 1950/2009.
[375] Rogers, Anna, Olga Kovaleva, and Anna Rumshisky. "A primer in BERTology: What we know about how BERT works." *Transactions of the Association for Computational Linguistics* 8 (2021): 842-866.
[376] Jiang, Zhiying, Matthew YR Yang, Mikhail Tsirlin, Raphael Tang, and Jimmy Lin. "Less is More: Parameter-Free Text Classification with Gzip." *arXiv preprint arXiv:2212.09410* (2022).
[377] Strubell, Emma, Ananya Ganesh, and Andrew McCallum. "Energy and policy considerations for deep learning in NLP." *arXiv preprint arXiv:1906.02243* (2019).
[378] Luitse, Dieuwertje, and Wiebke Denkena. "The great transformer: Examining the role of large language models in the political economy of AI." *Big Data & Society* 8, no. 2 (2021): 20539517211047734.
[379] Ferrara, Emilio. "Should chatgpt be biased? challenges and risks of bias in large language models." *arXiv preprint arXiv:2304.03738* (2023).
[380] Zhao, Wayne Xin, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min et al. "A survey of large language models." *arXiv preprint arXiv:2303.18223* (2023).
[381] Sheng, Emily, Kai-Wei Chang, Premkumar Natarajan, and Nanyun Peng. "The woman worked as a babysitter: On biases in language generation." *arXiv preprint arXiv:1909.01326* (2019).
[382] Bordia, Shikha, and Samuel R. Bowman. "Identifying and reducing gender bias in word-level language models." *arXiv preprint arXiv:1904.03035* (2019).
See also: Lu, Kaiji, Piotr Mardziel, Fangjing Wu, Preetam Amancharla, and Anupam Datta. "Gender bias in neural natural language processing." *Logic, Language, and Security: Essays Dedicated to Andre Scedrov on the Occasion of His 65th Birthday* (2020): 189-202.
See also: Caliskan, Aylin, Joanna J. Bryson, and Arvind Narayanan. "Semantics derived automatically from language corpora contain human-like biases." *Science* 356, no. 6334 (2017): 183-186.
See also generally: Bolukbasi, Tolga, Kai-Wei Chang, James Y. Zou, Venkatesh Saligrama, and Adam T. Kalai. "Man is to computer programmer as woman is to homemaker? debiasing word embeddings." *Advances in neural information processing systems* 29 (2016).

anti-Muslim,[383] and even political biases[384] from the underlying online text corpora. Professor Ferrara in his review of bias in large language models offers a convenient typology of biases, which are self-explanatory, namely: demographic biases,[385] cultural biases,[386] linguistic biases,[387] temporal biases,[388] confirmation biases,[389] and lastly, ideological and political biases.[390] It is important to briefly discuss what is meant by "bias" in the present work.[391] We adopt Kruglanski and Ajzen's characterization of bias in terms of preferences, namely that bias is "a subjectively-based preference for a given conclusion or inference over possible alternative conclusions."[392] Thus, in the context of the aforementioned typology of biases, for instance, a political ideological bias in LLMs can be defined as the model's preference for one political view over another political view due to the underlying lack of training data in the text representing the latter political

---

[383] Abid, Abubakar, Maheen Farooqi, and James Zou. "Persistent anti-muslim bias in large language models." In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, pp. 298-306. 2021.

[384] Rozado, David. "The political biases of chatgpt." *Social Sciences* 12, no. 3 (2023): 148.
See also: Rutinowski, Jérôme, Sven Franke, Jan Endendyk, Ina Dormuth, and Markus Pauly. "The Self-Perception and Political Biases of ChatGPT." *arXiv preprint arXiv:2304.07333* (2023).

[385] Kirk, Hannah Rose, Yennie Jun, Filippo Volpin, Haider Iqbal, Elias Benussi, Frederic Dreyer, Aleksandar Shtedritski, and Yuki Asano. "Bias out-of-the-box: An empirical analysis of intersectional occupational biases in popular generative language models." *Advances in neural information processing systems* 34 (2021): 2611-2624.

[386] Bordia, Shikha, and Samuel R. Bowman. "Identifying and reducing gender bias in word-level language models." *arXiv preprint arXiv:1904.03035* (2019).
See also: Blodgett, Su Lin, Solon Barocas, Hal Daumé III, and Hanna Wallach. "Language (technology) is power: A critical survey of" bias" in nlp." *arXiv preprint arXiv:2005.14050* (2020).

[387] Pires, Telmo, Eva Schlinger, and Dan Garrette. "How multilingual is multilingual BERT?." *arXiv preprint arXiv:1906.01502* (2019).
See also: Ruder, Sebastian, Ivan Vulić, and Anders Søgaard. "A survey of cross-lingual word embedding models." *Journal of Artificial Intelligence Research* 65 (2019): 569-631.

[388] Zellers, Rowan, Ari Holtzman, Hannah Rashkin, Yonatan Bisk, Ali Farhadi, Franziska Roesner, and Yejin Choi. "Defending against neural fake news." *Advances in neural information processing systems* 32 (2019).
See also: Lazaridou, Angeliki, Adhi Kuncoro, Elena Gribovskaya, Devang Agrawal, Adam Liska, Tayfun Terzi, Mai Gimenez et al. "Mind the gap: Assessing temporal generalization in neural language models." *Advances in Neural Information Processing Systems* 34 (2021): 29348-29363.

[389] Bolukbasi, Tolga, Kai-Wei Chang, James Y. Zou, Venkatesh Saligrama, and Adam T. Kalai. "Man is to computer programmer as woman is to homemaker? debiasing word embeddings." *Advances in neural information processing systems* 29 (2016).

[390] Hartmann, Jochen, Jasper Schwenzow, and Maximilian Witte. "The political ideology of conversational AI: Converging evidence on ChatGPT's pro-environmental, left-libertarian orientation." *arXiv preprint arXiv:2301.01768* (2023).

[391] For a critique of the existing definitions of bias in machine learning and NLP literature see generally: Blodgett, Su Lin, Solon Barocas, Hal Daumé III, and Hanna Wallach. "Language (technology) is power: A critical survey of" bias" in nlp." *arXiv preprint arXiv:2005.14050* (2020).

[392] Kruglanski, Arie W., and Icek Ajzen. "Bias and error in human judgment." *European Journal of Social Psychology* 13, no. 1 (1983): 1-44.

view. This implies that there exists an underlying set of political views, for example: conservative, liberal, communist, libertarian, etc. In a training corpus which contains 90% of textual data and discussion aligning with only liberal perspectives, it should not come as a surprise that a large language model trained on such a corpus will likely capture liberal-leaning patterns in text which will in turn lead to the generation of liberal-leaning outputs.[393] Bias in data can thus be analogized to voting, but instead of people, it is the data itself that "votes", i.e., in the previous example liberal perspective dominates over other perspectives simply due to the amount of liberal-aligning textual data that exists in the training corpus. It is important to point out that in addition to such "data biases" there are other factors that contribute to biases in large language models, such as the algorithms being used and the emphasis they tend to place on data, etc.[394]

In addition to the aforementioned data bias problems, another issue observed in large language models is the problem of so-called "hallucinations."[395] The problem of hallucinated output by large language models like ChatGPT is that it "gives the impression of being fluent and natural despite being unfaithful and nonsensical… and appears to be grounded in the real context provided, although it is actually hard to specify or verify the existence of such contexts."[396] Such hallucinations can be exploited to recover sensitive private information that was present in the training data, such as emails, etc.[397] From a technical perspective, the problem of hallucination itself stems from the utilization of standard likelihood maximization objective for training of the large language model[398] which results in bland and nonsensical text.[399]

To better illustrate the hallucination problem and its interaction with the underlying data bias problem, it is worth noting an illustrative tweet posted on May 28th, 2023, by

---

[393] Santurkar, Shibani, Esin Durmus, Faisal Ladhak, Cinoo Lee, Percy Liang, and Tatsunori Hashimoto. "Whose opinions do language models reflect?." *arXiv preprint arXiv:2303.17548* (2023).
[394] Ferrara, Emilio. "Should chatgpt be biased? challenges and risks of bias in large language models." *arXiv preprint arXiv:2304.03738* (2023).
[395] Maynez, Joshua, Shashi Narayan, Bernd Bohnet, and Ryan McDonald. "On faithfulness and factuality in abstractive summarization." *arXiv preprint arXiv:2005.00661* (2020).
[396] Ji, Ziwei, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. "Survey of hallucination in natural language generation." *ACM Computing Surveys* 55, no. 12 (2023): 1-38.
[397] Carlini, Nicholas, Florian Tramer, Eric Wallace, Matthew Jagielski, Ariel Herbert-Voss, Katherine Lee, Adam Roberts et al. "Extracting training data from large language models." In *30th USENIX Security Symposium (USENIX Security 21)*, pp. 2633-2650. 2021.
See also: Ji, Ziwei, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. "Survey of hallucination in natural language generation." *ACM Computing Surveys* 55, no. 12 (2023): 1-38.
[398] Welleck, Sean, Ilia Kulikov, Stephen Roller, Emily Dinan, Kyunghyun Cho, and Jason Weston. "Neural text generation with unlikelihood training." *arXiv preprint arXiv:1908.04319* (2019).
[399] Holtzman, Ari, Jan Buys, Li Du, Maxwell Forbes, and Yejin Choi. "The curious case of neural text degeneration." *arXiv preprint arXiv:1904.09751* (2019).

Nassim Taleb which shows his conversation with ChatGPT where he asked whether "Caratheodori met Disraeli."[400] Caratheodory is a surname of Greek origin, and Constantin Caratheodory, a 20th Century Greek mathematician, is perhaps the most famous representative of this family name, and indeed much text has been written about him and his mathematical work.[401] However, a lesser known fact is that Constantin Caratheodory's great uncle was Alexander Caratheodory, who was in fact a well-known statesman in the Ottoman Empire, achieving the rank of Pasha and becoming the Ottoman Minister of Foreign Affairs and thus, Alexander Caratheodory represented Turkey on various diplomatic occasions.

The answer that ChatGPT gave to Nassim Taleb's question whether "Caratheodori met Disraeli" however, was that "Caratheodori Pasha actively participated in the Congress [of Berlin in 1878] as a representative of *Greece*."[402] Nassim Taleb stated in his tweet that "ChatGPT likes the obvious, which is where the problem lies, missing BY CONSTRUCTION the ironies and nuances of history. Alexander Caratheodory was not the representative of Greece but the representative of "Turkey", rather the Ottoman Empire. He was the foreign minister of the Grande Porte."[403] This tweet illustrates that perhaps due to abundance of data on the Greek members of the Caratheodory family, and the relative lack of data on the Turkish and Ottoman members of the Caratheodory family, ChatGPT's output stated that Caratheodory was a representative of Greece in the Congress of Berlin rather than a representative of Turkey. Taleb's argument that this problem is in fact "by construction" is also accurate given the underlying problems of data biases, parameterization, optimization functions involved in the training, which likely lead to agglomerating and averaging of the informational content without regard to nuance. Thus, data biases combined with algorithmic choices in training result in the generation of inaccurate and false information.[404]

Large language models, despite not explicitly being designed to capture reasoning patterns, nevertheless, seem capable of doing so to a very limited extent.[405] For instance,

---

[400] Nassim Nicholas Taleb, 28 May 2023 tweet, 3:07 PM, available at: https://twitter.com/nntaleb/status/1662943708600303617

[401] Georgiadou, Maria, and Maria Georgiadou. *Constantin Carathéodory*. Springer-Verlag, 2003.

[402] Nassim Nicholas Taleb, 28 May 2023 tweet, 3:07 PM, available at: https://twitter.com/nntaleb/status/1662943708600303617

[403] *Ibid.*

[404] Parikh, Ankur P., Xuezhi Wang, Sebastian Gehrmann, Manaal Faruqui, Bhuwan Dhingra, Diyi Yang, and Dipanjan Das. "ToTTo: A controlled table-to-text generation dataset." *arXiv preprint arXiv:2004.14373* (2020).
See also: Zhou, Chunting, Graham Neubig, Jiatao Gu, Mona Diab, Paco Guzman, Luke Zettlemoyer, and Marjan Ghazvininejad. "Detecting hallucinated content in conditional neural sequence generation." *arXiv preprint arXiv:2011.02593* (2020).

[405] Brown, Tom, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D. Kaplan, Prafulla Dhariwal, Arvind Neelakantan et al. "Language models are few-shot learners." *Advances in neural information processing systems* 33 (2020): 1877-1901.

Yao et al. note that "GPT-3.5 series cannot do reasoning within formal, strict systems like math or first-order logic" but it can "do very well on mixture of logic and ambiguous statement [type of reasoning]."[406] It is worth noting that this statement was recently updated, stating that "connecting ChatGPT with Wolfram is an effective way of improving its formal/math reasoning ability."[407] In another paper, Wei et al. find that simple chain-of-thought reasoning prompting can improve "the ability of large language models to perform complex reasoning," which includes tasks such as, for instance, arithmetic, commonsense, and symbolic reasoning tasks.[408] More recently, it was discovered that by adding a simple "Let's think step by step" to a prompt leads to significant LLM performance gain on arithmetic, symbolic reasoning, and other logical reasoning tasks.[409] Takeshi et al. state that, "the versatility of this single prompt across very diverse reasoning tasks hints at untapped and understudied fundamental zero-shot capabilities of LLMs, suggesting high-level, multi-task broad cognitive capabilities may be extracted by simple prompting."[410] In other words, there seems to be some underlying capability of LLMs to capture reasoning patterns from the text and exhibit what we might categorize as reasoning.[411] However, it is worth pointing out that despite these alleged reasoning capabilities, recent work has argued that "despite the occasional flashes of analytical brilliance, GPT-4 at present is utterly incapable of reasoning."[412] And although this discussion is not settled, we adopt the more skeptical view of LLMs' reasoning capabilities, noting however, that textual patterns can capture linguistic cues representing the use of epistemological schemas of intelligibility.

The discussions regarding LLMs' reasoning capabilities fall under the topic of "emergent properties" of large language models, which are defined as properties or abilities of a model that are "not present in smaller models but [are] present in larger

---

For a great paper questioning ChatGPT's reasoning capabilities see:
Arkoudas, K. *GPT-4 Can't Reason*. *Preprints* 2023, 2023080148. https://doi.org/10.20944/preprints202308.0148.v2

[406] Fu, Yao, Hao Peng, and Tushar Khot. "How does gpt obtain its ability? tracing emergent abilities of language models to their sources." *Yao Fu's Notion* (2022).

[407] *Ibid.*

See also: Dong, Honghua, Jiayuan Mao, Tian Lin, Chong Wang, Lihong Li, and Denny Zhou. "Neural logic machines." *arXiv preprint arXiv:1904.11694* (2019).

[408] Wei, Jason, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V. Le, and Denny Zhou. "Chain-of-thought prompting elicits reasoning in large language models." *Advances in Neural Information Processing Systems* 35 (2022): 24824-24837.

[409] Kojima, Takeshi, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. "Large language models are zero-shot reasoners." *Advances in neural information processing systems* 35 (2022): 22199-22213.

[410] *Ibid.*

[411] For a recent in-depth discussion on defining and measuring intelligence in the context of artificial intelligence see: Chollet, François. "On the measure of intelligence." *arXiv preprint arXiv:1911.01547* (2019).

[412] Arkoudas, K. *GPT-4 Can't Reason*. *Preprints* 2023, 2023080148. https://doi.org/10.20944/preprints202308.0148.v2

models."[413] A key determinant of emergent properties is model scale, which is "measured by training compute and number of model parameters."[414] In other words, the fact that large language models contain billions of parameters leads to unforeseen and unintended capabilities, among them, possibilities of capturing reasoning patterns represented in the training data. In the words of Hahn and Goyal, "an ability is emergent in the sense that it arises without specialized training data or objectives, simply by scaling models and computation," and they conclude by stating that "the emergence of such behavior remains largely mysterious."[415] It can thus be hypothesized that if large language models can indeed capture reasoning and logical relationships in the training data, it is also foreseeable, given large language models' propensity towards capturing bias, that these LLMs could also capture bias of legal epistemic nature, just as these models capture other kinds of biases.[416]

3.5 Schemes of Intelligibility and Bias in LLMs

Having introduced Berthelot's schemes of intelligibility and their application to law, and having discussed the propensity of LLMs towards capturing bias and possibly even reasoning patterns from the training data, we can now combine the two arguments. Of course, Professor Samuel anticipates the exact problems that Berthelot's schemes of intelligibility pose for legal AI, noting that:

---

[413] Wei, Jason, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama et al. "Emergent abilities of large language models." *arXiv preprint arXiv:2206.07682* (2022).
[414] *Ibid.*
[415] Hahn, Michael, and Navin Goyal. "A theory of emergent in-context learning as implicit structure induction." *arXiv preprint arXiv:2303.07971* (2023).
[416] Note that there has also been research into linguistic "epistemological bias" in NLP, usually in the context of bias correction and neutralization. However, such studies define "epistemological bias" as bias "relating to linguistic features that subtly (often via presupposition) focus on the believability of a proposition", for example, when used in an example sentence, the "assertive 'stated' removes the bias introduces by 'claimed', which casts doubt on a statement." This sense of the term "epistemological bias" is distinct form the legal "epistemological bias" we propose in this essay, which we frame in terms of Berthelot's schemes of intelligibility. Therefore, to avoid confusion, the bias we propose in this essay is referred to as a legal epistemological bias, or schemas of intelligibly based bias.
See: Recasens, Marta, Cristian Danescu-Niculescu-Mizil, and Dan Jurafsky. "Linguistic models for analyzing and detecting biased language." In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1650-1659. 2013.
See also: Bordia, Shikha, and Samuel R. Bowman. "Identifying and reducing gender bias in word-level language models." *arXiv preprint arXiv:1904.03035* (2019)
See also: Madanagopal, Karthic, and James Caverlee. "Reinforced Sequence Training based Subjective Bias Correction." In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 2577-2590. 2023.

> "How is an AI programme going to handle these different scheme possibilities? One answer, of course, is not to have a single robot judge, but a college of them, different robots being programmed with different schemes of intelligibility and different paradigm orientations (holism versus individualism). Yet this would seemingly undermine part of the purpose of replacing human judges with a computerized judge supposedly free of human biases. It would undermine the idea of legal singularity."[417]

At this point in time, instead of creating six computer judges with different conceptual schemas, the current state of the art in NLP focuses on creating a single large language model trained on a training corpus made up of large amounts of text from the internet. This of course, raises issues of a bias of a novel, epistemological kind. Thus, from the perspective of legal reasoning, if schemes of intelligibility can be captured by large language models from the text, perhaps via emergent properties that we discussed earlier, then it follows that similar data biases might emerge with regards to what we can call the latent legal epistemological biases.

For instance, consider the following example: According to Samuel, "tort lawyers spend much time applying causal and actional methods in the analysis of case law problems."[418] Extending this to the problems of bias in large language models, we can thus make the prediction that a large language model trained on a corpus of only tort law cases would essentially be legally epistemologically biased in a conceptual schema sense towards applying only causal and actional schemes of intelligibility for generating its prompt responses. This assertion is also empirically testable, although that would require constructing a large language model, which is, unfortunately, prohibitively expensive at this point in time.

For future work, multiple things can be done. Firstly, it is clear that Berthelot's conceptual schemas of intelligibility can be used to develop human annotations of legal text corpora to develop a dataset of legal reasoning schemes. This has not been done yet, and to the best of our knowledge, no such dataset exists. As Professor Samuel rightly points out "these schemes are in need of much elaboration and explanation if their epistemological and methodological importance is to be fully appreciated."[419] Secondly, examining the schemes of intelligibility employed by large language models by examining their output seems to be the next logical step in our research agenda.

---

[417] Samuel, Geoffrey. "Is Legal Knowledge Regressing (Thanks to AI)?." *Series 2 Vol. 4 Amicus Curiae* (2023): 719.
[418] Samuel, Geoffrey. "Is law really a social science? A view from comparative law." *The Cambridge Law Journal* 67, no. 2 (2008)
[419] Samuel, Geoffrey. "Is law really a social science? A view from comparative law." *The Cambridge Law Journal* 67, no. 2 (2008): p 298

Thus, to go back to our initial point of inquiry: was Judge Padilla justified in using ChatGPT for his legal decision? And relatedly, can large language model-based AI be used to make legal decisions? As this essay tried to argue, the answer is a complicated one. Firstly, we discussed the nature of legal reasoning as being based on Berthelot's schemas of intelligibility. Secondly, we discussed the fact that ChatGPT and large language models generally tend to capture biases contained in the training data, in addition to the related problems of hallucination. They also seem to have emergent properties which are not well understood, however, these properties also seem to capture some semblance of rudimentary "reasoning."

Having established all this, we can predict that ChatGPT's answer to Judge Padilla, or any other legal question, will likely reflect the most dominant scheme of intelligibility in the underlying corpus given the necessary input, if such schemes of intelligibility are indeed capturable. Which schema of intelligibility ChatGPT adopts to a given factual pattern description is an empirical question that can be tested. As with the tort law corpus example, we can make a prediction that the dominant latent legal epistemological bias present in the underlying corpus will be reflected in the output.

A solution could be the specification of schema of intelligibly-based mode of reasoning to be applied in a given case – in a sense, a judge or a query provider could indeed specify which reasoning method to apply and thus explicitly constrain the model to provide only certain outputs. And we admit that this approach could indeed offer a viable solution to the problem. Nevertheless, this solution itself exposes a number of key questions regarding legal reasoning as such. Firstly, the ability to specify a mode of reasoning implies a knowledge of which mode of reasoning should be applied in a given legal case, and as we have tried to argue in this paper, legal epistemology and legal philosophy do not have a concrete answer as to why a certain schema of intelligibility should be preferred over another in a given case. Secondly, I believe that, following Samuel, such an approach could potentially "undermine legal singularity,"[420] or the idea of a singular legal solution, in that a different legal solution could emerge by asking the LLM to employ a different scheme of intelligibility to a given case. This in turn has the potential to expose the perspectivist reality of legal decision-making, thus laying bare the inherent subjectivity of judicial decisions and putting to question the centrality of legal objectivity in our societies as such.

[420] Samuel, Geoffrey. "Is Legal Knowledge Regressing (Thanks to AI)?." *Series 2 Vol. 4 Amicus Curiae* (2023): 719.

3.6 Conclusion

In this essay, motivated by Judge Padilla's use of ChatGPT for a legal decision, we attempted to tackle the question of LLM-based legal decision-making. This is an important question not because of the viability of AI decision-making, but because it poses important epistemological questions about legal reasoning. Our thesis is that while large language models are not explicitly designed to capture epistemological biases, they will still exhibit latent legal epistemological biases rooted in Berthelotian schema of intelligibility-based patterns of legal reasoning observed in the text.

To support this thesis, we discussed how large language models work, noting their propensity for bias and hallucination as a result of underlying training data and algorithmic issues, but also discussing their possible emergent properties, particularly relating to the possibility of capturing reasoning patterns in text. We also discussed legal reasoning, arguing that legal reasoning is essentially an application of different schemes of intelligibility to law and facts proposed by Berthelot.[421] Because bias in large language models can be seen as a consequence of the underlying data, we argued that certain schemas of intelligibility will be privileged over others depending on textual data upon which the large language model is trained on and we call this the latent legal epistemological bias. Thus, if the underlying textual data is mostly made up of legal decisions which employ the causal scheme, then a large language model trained on such a corpus will likely output decisions which also employ the causal scheme. Of course, all of this is predicated upon the assumption that large language models can indeed capture reasoning patterns in the text via emergent abilities or by some other means.

Future work is necessary to investigate Berthelot's schemas of intelligibility, particularly their linguistic cues in legal and other types of text. Furthermore, it is necessary to explore and empirically analyze how these schemes of intelligibility are used, exhibited, and manifested in large language models and whether certain schemas of intelligibility are indeed more dominant in legal texts and thus epistemologically privileged over others, resulting in the possible manifestation of latent legal epistemological biases.

[421] Berthelot's schemes of intelligibility are undoubtedly of fundamental importance to understanding legal reasoning, and we cannot do them proper justice in the span of the present work. We can only hope that Professor Berthelot's work will be translated into English sooner rather than later and that the linguistic cues associated with these schemes will be investigated further.

3.7 References

Abbott, Ryan, and Brinson S. Elliott. "Putting the Artificial Intelligence in Alternative Dispute Resolution: How AI Rules Will Become ADR Rules." *Amicus Curiae* 4, no. 3 (2023): 685-706.

Abid, Abubakar, Maheen Farooqi, and James Zou. "Persistent anti-muslim bias in large language models." In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, pp. 298-306. 2021.

Alexander, Larry, and Emily Sherwin. "Demystifying legal reasoning." (2008).

Arkoudas, K. *GPT-4 Can't Reason*. *Preprints* 2023, 2023080148. https://doi.org/10.20944/preprints202308.0148.v2

Armgardt, Matthias. "Leibniz as legal scholar." *Fundamina: A Journal of Legal History* 2014, no. si-1 (2014): 27-38.

Arrieta, Alejandro Barredo, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García et al. "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI." *Information fusion* 58 (2020): 82-115.

Arthur, Richard TW. *Leibniz*. John Wiley & Sons, 2014.

Berthelot, Jean-Michel. *Épistémologie des sciences sociales*. PUF, 2001.

Blodgett, Su Lin, Solon Barocas, Hal Daumé III, and Hanna Wallach. "Language (technology) is power: A critical survey of" bias" in nlp." *arXiv preprint arXiv:2005.14050* (2020).

Bolukbasi, Tolga, Kai-Wei Chang, James Y. Zou, Venkatesh Saligrama, and Adam T. Kalai. "Man is to computer programmer as woman is to homemaker? debiasing word embeddings." *Advances in neural information processing systems* 29 (2016).

Bordia, Shikha, and Samuel R. Bowman. "Identifying and reducing gender bias in word-level language models." *arXiv preprint arXiv:1904.03035* (2019).

Brown, Tom, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D. Kaplan, Prafulla Dhariwal, Arvind Neelakantan et al. "Language models are few-shot learners." *Advances in neural information processing systems* 33 (2020): 1877-1901.

Caliskan, Aylin, Joanna J. Bryson, and Arvind Narayanan. "Semantics derived automatically from language corpora contain human-like biases." *Science* 356, no. 6334 (2017): 183-186.

Carlini, Nicholas, Florian Tramer, Eric Wallace, Matthew Jagielski, Ariel Herbert-Voss, Katherine Lee, Adam Roberts et al. "Extracting training data from large language models." In *30th USENIX Security Symposium (USENIX Security 21)*, pp. 2633-2650. 2021.

Chollet, François. "On the measure of intelligence." *arXiv preprint arXiv:1911.01547* (2019).

Clark, Kevin, Urvashi Khandelwal, Omer Levy, and Christopher D. Manning. "What does bert look at? an analysis of bert's attention." *arXiv preprint arXiv:1906.04341* (2019).

Dascal, Marcelo. *Gottfried Wilhelm Leibniz: the art of controversies*. Edited by Quintín Racionero, and Adelino Cardoso. Springer Science+ Business Media BV, 2006.

de Montaigne, Michel. *Essays* (1595), translated by: JM Cohen, Penguin Books. (1957).

Deakin, Simon, and Christopher Markou, eds. *Is law computable?: critical perspectives on law and artificial intelligence*. Bloomsbury Publishing, 2020.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. "Bert: Pre-training of deep bidirectional transformers for language understanding." *arXiv preprint arXiv:1810.04805* (2018).

Dong, Honghua, Jiayuan Mao, Tian Lin, Chong Wang, Lihong Li, and Denny Zhou. "Neural logic machines." *arXiv preprint arXiv:1904.11694* (2019).

Dupuy, Jean-Pierre *Petite métaphysique des tsunamis*, Paris: Editions du Seuil 2005

Dworkin, Ronald. "Hard cases." *Harv. L. Rev.* 88 (1974): 1057.

Fedus, William, Barret Zoph, and Noam Shazeer. "Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity." *The Journal of Machine Learning Research* 23, no. 1 (2022): 5232-5270.

Ferrara, Emilio. "Should chatgpt be biased? challenges and risks of bias in large language models." *arXiv preprint arXiv:2304.03738* (2023).

Firth, John. "A synopsis of linguistic theory, 1930-1955." *Studies in linguistic analysis* (1957): 10-32.

Fordham, Margaret. "Loss of Chance-A Lost Opportunity." *Sing J. Legal Stud.* (2005): 204.

Fu, Yao, Hao Peng, and Tushar Khot. "How does gpt obtain its ability? tracing emergent abilities of language models to their sources." *Yao Fu's Notion* (2022).

Georgiadou, Maria, and Maria Georgiadou. *Constantin Carathéodory*. Springer-Verlag, 2003.

Gordley, James. *The jurists: a critical history*. OUP Oxford, 2013.

Hahn, Michael, and Navin Goyal. "A theory of emergent in-context learning as implicit structure induction." *arXiv preprint arXiv:2303.07971* (2023).

Haidt, Jonathan. *The righteous mind: Why good people are divided by politics and religion*. Vintage, 2012.

Hales, Steven D., and Rex Welshon. *Nietzsche's perspectivism*. University of Illinois Press, 2000.

Hartmann, Jochen, Jasper Schwenzow, and Maximilian Witte. "The political ideology of conversational AI: Converging evidence on ChatGPT's pro-environmental, left-libertarian orientation." *arXiv preprint arXiv:2301.01768* (2023).

Hilliard, Airlie, Nigel Guenole, and Franziska Leutner. "Robots are judging me: Perceived fairness of algorithmic recruitment tools." *Frontiers in Psychology* 13 (2022): 940456.

Hoeflich, Michael H. "Law & geometry: Legal science from Leibniz to Langdell." *American Journal of Legal History* 30, no. 2 (1986): 95-121.

Holtzman, Ari, Jan Buys, Li Du, Maxwell Forbes, and Yejin Choi. "The curious case of neural text degeneration." *arXiv preprint arXiv:1904.09751* (2019).

Hume, David. *A treatise of human nature*. Clarendon Press, 1896.

Ji, Ziwei, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. "Survey of hallucination in natural language generation." *ACM Computing Surveys* 55, no. 12 (2023): 1-38.

Jiang, Zhiying, Matthew YR Yang, Mikhail Tsirlin, Raphael Tang, and Jimmy Lin. "Less is More: Parameter-Free Text Classification with Gzip." *arXiv preprint arXiv:2212.09410* (2022).

Jones, Steven E. *Against technology: From the Luddites to neo-Luddism*. Routledge, 2013.

Jurafsky, Daniel, and James H. Martin. "Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition." 2023 ed.

Kelsen, Hans. *Pure theory of law*. Univ of California Press, 1967.

Keynes, John Maynard. *They Economic Possibilities of Our Grandchildren* (1930)

Kirk, Hannah Rose, Yennie Jun, Filippo Volpin, Haider Iqbal, Elias Benussi, Frederic Dreyer, Aleksandar Shtedritski, and Yuki Asano. "Bias out-of-the-box: An empirical analysis of intersectional occupational biases in popular generative language models." *Advances in neural information processing systems* 34 (2021): 2611-2624.

Kojima, Takeshi, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. "Large language models are zero-shot reasoners." *Advances in neural information processing systems* 35 (2022): 22199-22213.

Kruglanski, Arie W., and Icek Ajzen. "Bias and error in human judgment." *European Journal of Social Psychology* 13, no. 1 (1983): 1-44.

Lazaridou, Angeliki, Adhi Kuncoro, Elena Gribovskaya, Devang Agrawal, Adam Liska, Tayfun Terzi, Mai Gimenez et al. "Mind the gap: Assessing temporal generalization in neural language models." *Advances in Neural Information Processing Systems* 34 (2021): 29348-29363.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." *nature* 521, no. 7553 (2015): 436-444.

Leibniz, Gottfried Wilhelm. "Elements of Natural Law." In *Philosophical papers and letters*, pp. 131-138. Springer Netherlands, 1989. 2nd edition Translated by: Leroy E. Loemker

Lu, Kaiji, Piotr Mardziel, Fangjing Wu, Preetam Amancharla, and Anupam Datta. "Gender bias in neural natural language processing." *Logic, Language, and Security: Essays Dedicated to Andre Scedrov on the Occasion of His 65th Birthday* (2020): 189-202.

Luitse, Dieuwertje, and Wiebke Denkena. "The great transformer: Examining the role of large language models in the political economy of AI." *Big Data & Society* 8, no. 2 (2021): 20539517211047734.

Madanagopal, Karthic, and James Caverlee. "Reinforced Sequence Training based Subjective Bias Correction." In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 2577-2590. 2023.

Martin, Andrew D., and Kevin M. Quinn. "Dynamic ideal point estimation via Markov chain Monte Carlo for the US Supreme Court, 1953–1999." *Political analysis* 10, no. 2 (2002): 134-153.

Maynez, Joshua, Shashi Narayan, Bernd Bohnet, and Ryan McDonald. "On faithfulness and factuality in abstractive summarization." *arXiv preprint arXiv:2005.00661* (2020).

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. "Distributed representations of words and phrases and their compositionality." *Advances in neural information processing systems* 26 (2013).

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. "Efficient estimation of word representations in vector space." *arXiv preprint arXiv:1301.3781* (2013).

Morison, John, and Adam Harkens. "Re-engineering justice? Robot judges, computerised courts and (semi) automated legal decision-making." *Legal Studies* 39, no. 4 (2019): 618-635.

Nietzsche, Freidrich, *Will to Power* (1901/1968), Vintage, translated by Kaufman

Parikh, Ankur P., Xuezhi Wang, Sebastian Gehrmann, Manaal Faruqui, Bhuwan Dhingra, Diyi Yang, and Dipanjan Das. "ToTTo: A controlled table-to-text generation dataset." *arXiv preprint arXiv:2004.14373* (2020).

Parikh, Purvish M., Dinesh M. Shah, and Kairav P. Parikh. "Judge Juan Manuel Padilla Garcia, ChatGPT, and a controversial medicolegal milestone." *Indian Journal of Medical Sciences* 75, no. 1 (2023): 3-8.

Peng, Baolin, Michel Galley, Pengcheng He, Hao Cheng, Yujia Xie, Yu Hu, Qiuyuan Huang et al. "Check your facts and try again: Improving large language models with external knowledge and automated feedback." *arXiv preprint arXiv:2302.12813* (2023).

Pires, Telmo, Eva Schlinger, and Dan Garrette. "How multilingual is multilingual BERT?." *arXiv preprint arXiv:1906.01502* (2019).

Portugal, Ivens, Paulo Alencar, and Donald Cowan. "The use of machine learning algorithms in recommender systems: A systematic review." *Expert Systems with Applications* 97 (2018): 205-227.

Propp, Vladimir. *Morphology of the Folktale*. University of Texas Press, 1968.

Propp, Vladimir. *Theory and history of folklore*. Vol. 5. U of Minnesota Press, 1985.

Radford, Alec, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. "Improving language understanding by generative pre-training." (2018).

Recasens, Marta, Cristian Danescu-Niculescu-Mizil, and Dan Jurafsky. "Linguistic models for analyzing and detecting biased language." In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1650-1659. 2013.

Reinert, Hugo, and Erik S. Reinert. "Creative destruction in economics: Nietzsche, Sombart, schumpeter." *Friedrich Nietzsche (1844–1900) Economy and Society* (2006): 55-85.

Ren, Xiaozhe, Pingyi Zhou, Xinfan Meng, Xinjing Huang, Yadao Wang, Weichao Wang, Pengfei Li et al. "PanGu-{\Sigma}: Towards Trillion Parameter Language Model with Sparse Heterogeneous Computing." *arXiv preprint arXiv:2303.10845* (2023).

Rogers Anna, Olga Kovaleva, and Anna Rumshisky. "A primer in BERTology: What we know about how BERT works." *Transactions of the Association for Computational Linguistics* 8 (2021): 842-866.

Rozado, David. "The political biases of chatgpt." *Social Sciences* 12, no. 3 (2023): 148.

Ruder, Sebastian, Ivan Vulić, and Anders Søgaard. "A survey of cross-lingual word embedding models." *Journal of Artificial Intelligence Research* 65 (2019): 569-631.

Rutinowski, Jérôme, Sven Franke, Jan Endendyk, Ina Dormuth, and Markus Pauly. "The Self-Perception and Political Biases of ChatGPT." *arXiv preprint arXiv:2304.07333* (2023).

Samuel, Geoffrey. "Can legal reasoning be demystified?." *Legal Studies* 29, no. 2 (2009): 181-210.

Samuel, Geoffrey. "Is law really a social science? A view from comparative law." *The Cambridge Law Journal* 67, no. 2 (2008): 288-321.

Samuel, Geoffrey. "Is Legal Knowledge Regressing (Thanks to AI)?." *Amicus Curiae* 4, no. 3 (2023): 719-750.

Samuel, Geoffrey. "Taking Methods Seriously (Part One)." *J. Comp. L.* 2 (2007): 94.

Samuel, Geoffrey. "The challenge of artificial intelligence: can Roman law help us discover whether law is a system of rules?." *Legal Studies* 11, no. 1 (1991): 24-46.

Samuel, Geoffrey. *Epistemology and method in law*. Routledge, 2016.

Samuel, Geoffrey. *Rethinking legal reasoning*. Edward Elgar Publishing, 2018.

Santurkar, Shibani, Esin Durmus, Faisal Ladhak, Cinoo Lee, Percy Liang, and Tatsunori Hashimoto. "Whose opinions do language models reflect?." *arXiv preprint arXiv:2303.17548* (2023).

Schopenhauer, Arthur*: 'The World as Will and Representation': Volume 1* (1818). Cambridge University Press, 2010 ed. Trans: Judith Norman, Alistair Welchman and Christopher Janaway

Sheng, Emily, Kai-Wei Chang, Premkumar Natarajan, and Nanyun Peng. "The woman worked as a babysitter: On biases in language generation." *arXiv preprint arXiv:1909.01326* (2019).

Sourdin, Tania. "Judge v Robot?: Artificial intelligence and judicial decision-making." *University of New South Wales Law Journal, The* 41, no. 4 (2018): 1114-1133.

Strubell, Emma, Ananya Ganesh, and Andrew McCallum. "Energy and policy considerations for deep learning in NLP." *arXiv preprint arXiv:1906.02243* (2019).

Susskind, Richard. "Online courts and the future of justice." (2019).

Susskind, Richard. *Expert systems in law*. Oxford University Press, Inc., 1987.

Susskind, Richard. *Tomorrow's lawyers: An introduction to your future*. Oxford University Press, 2013.

Turing, Alan M. *Computing machinery and intelligence*. Springer Netherlands, 1950/2009.

Vamathevan, Jessica, Dominic Clark, Paul Czodrowski, Ian Dunham, Edgardo Ferran, George Lee, Bin Li et al. "Applications of machine learning in drug discovery and development." *Nature reviews Drug discovery* 18, no. 6 (2019): 463-477.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. "Attention is all you need." *Advances in neural information processing systems* 30 (2017).

Volokh, Eugene. "Chief justice robots." *Duke LJ* 68 (2018): 1135.

von Savigny, Friedrich Carl. *Of the Vocation of Our Age for Legislation and Jurisprudence*. Littlewood and Company, 1831. Translated by: Abraham Hayward, p 38-39

Wang, Nu. "“Black Box Justice”: Robot Judges and AI-based Judgment Processes in China’s Court System." In *2020 IEEE International Symposium on Technology and Society (ISTAS)*, pp. 58-65. IEEE, 2020.

Wei, Jason, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V. Le, and Denny Zhou. "Chain-of-thought prompting elicits reasoning in large language models." *Advances in Neural Information Processing Systems* 35 (2022): 24824-24837.

Wei, Jason, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama et al. "Emergent abilities of large language models." *arXiv preprint arXiv:2206.07682* (2022).

Welleck, Sean, Ilia Kulikov, Stephen Roller, Emily Dinan, Kyunghyun Cho, and Jason Weston. "Neural text generation with unlikelihood training." *arXiv preprint arXiv:1908.04319* (2019).

Xu, Zichun. "Human Judges in the era of artificial intelligence: challenges and opportunities." *Applied Artificial Intelligence* 36, no. 1 (2022): 2013652.

Zellers, Rowan, Ari Holtzman, Hannah Rashkin, Yonatan Bisk, Ali Farhadi, Franziska Roesner, and Yejin Choi. "Defending against neural fake news." *Advances in neural information processing systems* 32 (2019).

Zhao, Wayne Xin, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min et al. "A survey of large language models." *arXiv preprint arXiv:2303.18223* (2023).

Zhou, Chunting, Graham Neubig, Jiatao Gu, Mona Diab, Paco Guzman, Luke Zettlemoyer, and Marjan Ghazvininejad. "Detecting hallucinated content in conditional neural sequence generation." *arXiv preprint arXiv:2011.02593* (2020).

Žižek, Slavoj. *Event: Philosophy in transit*. Penguin UK, 2014.

Zizek, Slavoj. *The sublime object of ideology*. Verso Books, 2009

4 Conclusions

> *Mis palabras, que ahora son el presente,*
> *serán apenas la memoria de un sueño.*
>
> [My words, which are now your present,
> will one day be but the vaguest memory of a dream.]
>
> Jorge Luis Borges, *August 25, 1983* (1983)[422]

The present work, titled "*Essays on Law and Natural Language Processing*" explored the uses of Natural Language Processing (NLP) in law.

In the first essay, we proposed the application of binary text classification methodology to the question of predicting the outcome of the Supreme Court's certiorari petition decisions. Modeling the problem of certiorari petition outcome prediction as a text-based binary text classification task provides important benefits in that it enables us to explore the legal issues in the petitions that the text represents, and thus, allows us to examine the role that text potentially plays in determining certiorari outcomes. We employ topic-based features to capture legal issues in the petitions under the fourteenth amendment, however, other representational methods are of course possible and viable. Another notable benefit of the binary text classification approach is its likely extendibility to similar types of legal petition and motion processes irrespective of the underlying language of the textual data or the jurisdiction in question. However, the methodology is of course not perfect, especially in light of the large class imbalance issues that are inherent in the certiorari data-generating process. Although we focused solely on examining the text's role in determining certiorari outcomes, the feature space of the classification task can be extended with other data beyond text, for instance, other possible cues[423] such as the political leanings of the judges presiding at the time of certiorari review, etc.

In the second essay, we proposed the application of word embedding-based synchronic lexical semantic change detection methodology to investigate a question in corporate governance regarding corporate honesty on Environmental, Social and Governance (ESG) issues. We explore the question by separating corporate discourse into

[422] Borges, Jorge Luis. *Collected fictions*. Penguin, 1999. From a collection of short stories titled *Shakespeare's Memory*
[423] Tanenhaus, Joseph, et al. "The Supreme Court's certiorari jurisdiction: Cue theory." *Judicial decision making* 111 (1963).

distinct "corporate discourse channels,"[424] thus assuming the existence of "candid" and "official" corporate discourse channels. Subsequently, we propose defining corporate honesty as measurable inconsistency in language use across these corporate discourse channels. To measure this inconsistency on the issue of ESG, we employ Hamilton et al.[425] and Schlechtweg et al.[426] lexical semantic change detection techniques. These techniques allow us to examine the modalities of semantic change synchronically in terms of distinct domains using cosine distance measures. Applying synchronic lexical semantic change detection techniques to the task of measuring corporate honesty on the issue of ESG yields results indicative of consistency in corporate language across the candid and official discourse channels. Alternatively, the results can also be interpreted as a confirmation of the "likely influences of similar economic factors... in different discourse channels,"[427] in addition to confirming the idea of "language management"[428] inherent in corporate communication, which is why we do not see any notable difference in language use on ESG or other corporate issues, suggesting that our separation of corporate communication into these specific discourse channels is not reflective of candid or official corporate communication, but rather indicates that corporate communication is identical regardless of the channel employed. Although there are a number of limitations particularly when it comes to the data, the underlying methodology of synchronic lexical semantic change detection is indeed an effective technique for exploring questions of distinct language use across various communicative and discursive domains.

In the third essay, we discussed the legal epistemological questions surrounding the problem of the use of large language models for legal decision-making. First, we discussed the illustrative example of Judge Padilla, who recently used ChatGPT for deciding a legal case before him. We subsequently discussed theories of legal reasoning, particularly focusing on Berthelot's schemas of intelligibility-based model of legal decision-making,[429] and we demonstrated that these schemas of intelligibility can be reflected in the text of legal decisions. We subsequently discussed large language models and their propensity to capture patterns and also bias in the underlying training data, in addition to the alleged reasoning capabilities of LLMs. Thus, the argument we made is that although large language models are not explicitly designed to capture reasoning

---

[424] Li, Feng. "Textual analysis of corporate disclosures: A survey of the literature." *Journal of accounting literature* 29 (2010): 143.
[425] Hamilton, William L., Jure Leskovec, and Dan Jurafsky. "Diachronic word embeddings reveal statistical laws of semantic change." *arXiv preprint arXiv:1605.09096* (2016).
[426] Schlechtweg, Dominik, Anna Hätty, Marco Del Tredici, and Sabine Schulte im Walde. "A wind of change: Detecting and evaluating lexical semantic change across times and domains." *arXiv preprint arXiv:1906.02979* (2019).
[427] Li, Feng. "Textual analysis of corporate disclosures: A survey of the literature." *Journal of accounting literature* 29 (2010): 143.
[428] Bargiela, Francesca, and Sandra J. Harris. *Managing language: The discourse of corporate meetings*. Vol. 44. John Benjamins Publishing, 1997.
[429] Berthelot, Jean-Michel. *Épistémologie des sciences sociales*. PUF, 2001.
See also: Samuel, Geoffrey. *Epistemology and method in law*. Routledge, 2016.

patterns, they will likely exhibit latent legal epistemological biases rooted in Berthelotian schemas of intelligibility-based patterns of legal reasoning as reflected in the training text data. Accordingly, this can limit the decision-making capabilities of these systems as they will inherently privilege the most dominant schemas of intelligibility that are present in the underlying data. We concluded with proposals to explore the linguistic cues of Berthelotian schemes of intelligibility and the need to closely examine the output of large language models when presented with legal situations to confirm whether LLMs can indeed employ Berthelot's schemes of intelligibility and thus possibly exhibit latent legal epistemological biases.

To conclude, this present work explored and applied various NLP techniques to legal questions. We started with binary text classification and topic modeling for certiorari outcome prediction, subsequently proceeding to word embedding based lexical semantic change detection in the context of measuring corporate honesty, and concluding with a discussion of large language models and their pattern-detecting capabilities and biases in the context of their use for legal decision-making and their possibility of exhibiting latent legal epistemological biases learned from the patterns in the underlying training data. A common theme that emerges throughout the present work is the underlying complexity of the data generating processes and consequently the problems involved in using NLP techniques with language generated in the various legal domains. Firstly, in the case of certiorari petitions, this was demonstrated by the complexity of the judicial decision-making space in that the actual certiorari decisions might not be guided by the text of said petitions. Secondly, in the case of corporate discourse, this was demonstrated by the corporate actors' sophistication and their ability to manage their language. Thirdly, in the case of the discussion on the possibilities of large language model-based legal decision-making, this was demonstrated by the fundamentally perspectivist nature of legal reasoning and the epistemic model of legal reasoning and legal decision-making provided by Berthelot's theory of schemas of intelligibility. Therefore, careful attention must be paid to the underlying textual data-generating processes involved and despite the benefits that NLP methods present, a general skeptical attitude towards NLP applications in law must nevertheless be adopted.

5 Bibliography

Abbott, Ryan, and Brinson S. Elliott. "Putting the Artificial Intelligence in Alternative Dispute Resolution: How AI Rules Will Become ADR Rules." *Amicus Curiae* 4, no. 3 (2023): 685-706.

Abid, Abubakar, Maheen Farooqi, and James Zou. "Persistent anti-muslim bias in large language models." In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, pp. 298-306. 2021.

Aggarwal, Charu C., and Charu C. Aggarwal. Machine learning for text: An introduction. Springer International Publishing, 2018.

Aletras, Nikolaos, Dimitrios Tsarapatsanis, Daniel Preoţiuc-Pietro, and Vasileios Lampos. "Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective." PeerJ computer science 2 (2016): e93.

Alexander, Larry, and Emily Sherwin. "Demystifying legal reasoning." (2008).

Amar, Akhil Reed. "The bill of rights and the fourteenth amendment." Yale LJ 101 (1991).

Amernic, Joel, and Russell Craig. *CEO-speak: The language of corporate leadership*. McGill-Queen's Press-MQUP, 2006.

Arkoudas, K. *GPT-4 Can't Reason*. *Preprints* 2023, 2023080148. https://doi.org/10.20944/preprints202308.0148.v2

Armgardt, Matthias. "Leibniz as legal scholar." *Fundamina: A Journal of Legal History* 2014, no. si-1 (2014): 27-38.

Arrieta, Alejandro Barredo, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García et al. "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI." *Information fusion* 58 (2020): 82-115.

Arthur, Richard TW. *Leibniz*. John Wiley & Sons, 2014.

Arun, Rajkumar, Venkatasubramaniyan Suresh, C. E. Veni Madhavan, and M. N. Narasimha Murthy. "On finding the natural number of topics with latent dirichlet allocation: Some observations." In *Advances in Knowledge Discovery and Data Mining: 14th Pacific-Asia Conference, PAKDD 2010, Hyderabad, India, June 21-24, 2010. Proceedings. Part I 14*, pp. 391-402. Springer Berlin Heidelberg, 2010.

Bamman, David, and Noah Smith. "Contextualized sarcasm detection on twitter." proceedings of the international AAAI conference on web and social media. Vol. 9. No. 1. 2015.

Bamman, David, Jacob Eisenstein, and Tyler Schnoebelen. "Gender identity and lexical variation in social media." Journal of Sociolinguistics 18, no. 2 (2014): 135-160.

Bargiela, Francesca, and Sandra J. Harris. *Managing language: The discourse of corporate meetings*. Vol. 44. John Benjamins Publishing, 1997.

Basu, Chumki, Haym Hirsh, and William Cohen. "Recommendation as classification: Using social and content-based information in recommendation." Aaai/iaai. 1998.

Berk, Jonathan, and Jules H. Van Binsbergen. "The impact of impact investing." (2021).

Berthelot, Jean-Michel. *Épistémologie des sciences sociales*. PUF, 2001.

Bird, Steven, Ewan Klein, and Edward Loper. Natural language processing with Python: analyzing text with the natural language toolkit. " O'Reilly Media, Inc.", 2009.

Blair, J. Pete, Torsten O. Reimer, and Timothy R. Levine. "The role of consistency in detecting deception: The superiority of correspondence over coherence." *Communication Studies* 69, no. 5 (2018): 483-498.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent dirichlet allocation." *Journal of machine Learning research* 3, no. Jan (2003)

Blodgett, Su Lin, Solon Barocas, Hal Daumé III, and Hanna Wallach. "Language (technology) is power: A critical survey of" bias" in nlp." *arXiv preprint arXiv:2005.14050* (2020).

Boehm, Paul D., Gregory S. Douglas, William A. Burns, Paul J. Mankiewicz, David S. Page, and A. Edward Bence. "Application of petroleum hydrocarbon chemical fingerprinting and allocation techniques after the Exxon Valdez oil spill." *Marine Pollution Bulletin* 34, no. 8 (1997): 599-613.

Bolukbasi, Tolga, Kai-Wei Chang, James Y. Zou, Venkatesh Saligrama, and Adam T. Kalai. "Man is to computer programmer as woman is to homemaker? debiasing word embeddings." *Advances in neural information processing systems* 29 (2016).

Bordia, Shikha, and Samuel R. Bowman. "Identifying and reducing gender bias in word-level language models." *arXiv preprint arXiv:1904.03035* (2019).

Borges, Jorge Luis. *Collected fictions*. Penguin, 1999

Bottou, Léon, Corinna Cortes, John S. Denker, Harris Drucker, Isabelle Guyon, Larry D. Jackel, Yann LeCun et al. "Comparison of classifier methods: a case study in handwritten digit recognition." In Proceedings of the 12th IAPR International Conference on Pattern Recognition, Vol. 3-Conference C: Signal Processing (Cat. No. 94CH3440-5), vol. 2, pp. 77-82. IEEE, 1994.

Boutsidis, Christos, and Efstratios Gallopoulos. "SVD based initialization: A head start for nonnegative matrix factorization." Pattern recognition 41, no. 4 (2008): 1350-1362.

Brandt, Jakob, and Emil Lanzén. "A comparative review of SMOTE and ADASYN in imbalanced data classification." (2021).

Brown, Stephen V., and Jennifer Wu Tucker. "Large‐sample evidence on firms' year‐over‐year MD&A modifications." *Journal of Accounting Research* 49, no. 2 (2011): 309-346.

Brown, Tom, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D. Kaplan, Prafulla Dhariwal, Arvind Neelakantan et al. "Language models are few-shot learners." *Advances in neural information processing systems* 33 (2020): 1877-1901.

Brownlee, Jason. *Imbalanced classification with Python: better metrics, balance skewed classes, cost-sensitive learning*. 2020.

Brunet, Jean-Philippe, Pablo Tamayo, Todd R. Golub, and Jill P. Mesirov. "Metagenes and molecular pattern discovery using matrix factorization." Proceedings of the national academy of sciences 101, no. 12 (2004): 4164-4169.

Buckland, Michael, and Fredric Gey. "The relationship between recall and precision." *Journal of the American society for information science* 45, no. 1 (1994): 12-19.

Burman, Prabir. "A comparative study of ordinary cross-validation, v-fold cross-validation and the repeated learning-testing methods." Biometrika 76, no. 3 (1989): 503-514.

Bushnell, James B. "The economics of carbon offsets." In *The design and implementation of US climate policy*, pp. 197-209. University of Chicago Press, 2011.

Caldeira, Gregory A., and John R. Wright. "Amici curiae before the supreme court: who participates, when, and how much?." The Journal of Politics 52.3 (1990): 782-806.

Caldeira, Gregory A., and John R. Wright. "Organized interests and agenda setting in the US Supreme Court." American Political Science Review 82.4 (1988): 1109-1127.

Caliskan, Aylin, Joanna J. Bryson, and Arvind Narayanan. "Semantics derived automatically from language corpora contain human-like biases." *Science* 356, no. 6334 (2017): 183-186.

Camilleri, Mark Anthony. "Theoretical insights on integrated reporting: The inclusion of non-financial capitals in corporate disclosures." *Corporate Communications: An International Journal* 23, no. 4 (2018): 567-581.

Campbell, John L., Hsinchun Chen, Dan S. Dhaliwal, Hsin-min Lu, and Logan B. Steele. "The information content of mandatory risk factor disclosures in corporate filings." *Review of Accounting Studies* 19 (2014): 396-455.

Carlini, Nicholas, Florian Tramer, Eric Wallace, Matthew Jagielski, Ariel Herbert-Voss, Katherine Lee, Adam Roberts et al. "Extracting training data from large language models." In *30th USENIX Security Symposium (USENIX Security 21)*, pp. 2633-2650. 2021.

Cavnar, William B., and John M. Trenkle. "N-gram-based text categorization." Proceedings of SDAIR-94, 3rd annual symposium on document analysis and information retrieval. Vol. 161175. 1994.

Cazier, Richard A., and Ray J. Pfeiffer. "Why are 10-K filings so long?." *Accounting Horizons* 30, no. 1 (2016): 1-21.

Chang, Yin-Wen, and Chih-Jen Lin. "Feature ranking using linear SVM." In Causation and prediction challenge, pp. 53-64. PMLR, 2008.

Chawla, Nitesh V., Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. "SMOTE: synthetic minority over-sampling technique." Journal of artificial intelligence research 16 (2002): 321-357.

Chen, Shuping, Stephan Hollander, and Kelvin Law. "In search of interaction." *Available at SSRN 2449341* (2016).

Chen, Tianqi, and Carlos Guestrin. "Xgboost: A scalable tree boosting system." In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pp. 785-794. 2016.

Chichilnisky, Graciela, and Geoffrey Heal. "Who should abate carbon emissions?: An international viewpoint." *Economics Letters* 44, no. 4 (1994): 443-449.

Chiu, Tiffany, Feiqi Huang, Yue Liu, and Miklos A. Vasarhelyi. "The impact of non-timely 10-Q filings and audit firm size on audit fees." *Managerial Auditing Journal* 33, no. 5 (2018): 503-516.

Chollet, François. "On the measure of intelligence." *arXiv preprint arXiv:1911.01547* (2019).

Christensen, Dane M., George Serafeim, and Anywhere Sikochi. "Why is corporate virtue in the eye of the beholder? The case of ESG ratings." *The Accounting Review* 97, no. 1 (2022): 147-175.

Cicero, Marcus Tullius, *On Divination* (44 BC)

Clark, Kevin, Urvashi Khandelwal, Omer Levy, and Christopher D. Manning. "What does bert look at? an analysis of bert's attention." *arXiv preprint arXiv:1906.04341* (2019).

Cortes, Corinna, and Vladimir Vapnik. "Support-vector networks." Machine learning 20 (1995): 273-297.

Csutora, Maria, and Gabor Harangozo. "Twenty years of carbon accounting and auditing–a review and outlook." *Society and Economy* 39, no. 4 (2017): 459-480.

Damodaran, Aswath. *Corporate finance*. Hoboken: Wiley, 1996.

Danescu-Niculescu-Mizil, Cristian, Lillian Lee, Bo Pang, and Jon Kleinberg. "Echoes of power: Language effects and power differences in social interaction." In *Proceedings of the 21st international conference on World Wide Web*, pp. 699-708. 2012.

Dascal, Marcelo. *Gottfried Wilhelm Leibniz: the art of controversies*. Edited by Quintín Racionero, and Adelino Cardoso. Springer Science+ Business Media BV, 2006.

de Montaigne, Michel. *Essays* (1595), translated by: JM Cohen, Penguin Books. (1957).

Deakin, Simon, and Christopher Markou, eds. *Is law computable?: critical perspectives on law and artificial intelligence*. Bloomsbury Publishing, 2020.

Dechezleprêtre, Antoine, Adrien Fabre, Tobias Kruse, Bluebery Planterose, Ana Sanchez Chico, and Stefanie Stantcheva. *Fighting climate change: International attitudes toward climate policies*. No. w30265. National Bureau of Economic Research, 2022.

Delmas, Magali A., and Vanessa Cuerel Burbano. "The drivers of greenwashing." *California management review* 54, no. 1 (2011): 64-87.

Devins, Neal, and Lawrence Baum. "Split definitive: How party polarization turned the Supreme Court into a partisan court." The Supreme Court Review 2016, no. 1 (2017): 301-365.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. "Bert: Pre-training of deep bidirectional transformers for language understanding." *arXiv preprint arXiv:1810.04805* (2018).

Dong, Honghua, Jiayuan Mao, Tian Lin, Chong Wang, Lihong Li, and Denny Zhou. "Neural logic machines." *arXiv preprint arXiv:1904.11694* (2019).

Dorfleitner, Gregor, Gerhard Halbritter, and Mai Nguyen. "Measuring the level and risk of corporate responsibility–An empirical comparison of different ESG rating approaches." Journal of Asset Management 16 (2015): 450-466.

Dupuy, Jean-Pierre *Petite métaphysique des tsunamis*, Paris: Editions du Seuil 2005

Dworkin, Ronald. "Hard cases." *Harv. L. Rev.* 88 (1974): 1057.

Fedus, William, Barret Zoph, and Noam Shazeer. "Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity." *The Journal of Machine Learning Research* 23, no. 1 (2022): 5232-5270.

Feely, Alan J., and Anne‐Wil Harzing. "Language management in multinational companies." *Cross Cultural Management: an international journal* 10, no. 2 (2003): 37-52.

Feldman, Adam, and Alexander Kappner. "Finding certainty in cert: An empirical analysis of the factors involved in Supreme Court certiorari decisions from 2001-2015." Vill. L. Rev. 61 (2016): 795.

Fernández, Alberto, Salvador Garcia, Francisco Herrera, and Nitesh V. Chawla. "SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary." Journal of artificial intelligence research 61 (2018): 863-905.

Fernández, Alberto, Salvador García, Mikel Galar, Ronaldo C. Prati, Bartosz Krawczyk, and Francisco Herrera. Learning from imbalanced data sets. Vol. 10. Cham: Springer, 2018.

Ferrara, Emilio. "Should chatgpt be biased? challenges and risks of bias in large language models." *arXiv preprint arXiv:2304.03738* (2023).

Ferrari, Alessio, Beatrice Donati, and Stefania Gnesi. "Detecting domain-specific ambiguities: an NLP approach based on Wikipedia crawling and word embeddings." In *2017 IEEE 25th International Requirements Engineering Conference Workshops (REW)*, pp. 393-399. IEEE, 2017.

Firth, John. "A synopsis of linguistic theory, 1930-1955." *Studies in linguistic analysis* (1957): 10-32.

Fordham, Margaret. "Loss of Chance-A Lost Opportunity." *Sing J. Legal Stud.* (2005): 204.

Fu, Yao, Hao Peng, and Tushar Khot. "How does gpt obtain its ability? tracing emergent abilities of language models to their sources." *Yao Fu's Notion* (2022).

Galasinski, Dariusz. *The language of deception: A discourse analytical study*. Sage Publications, 2000.

Georgiadou, Maria, and Maria Georgiadou. *Constantin Carathéodory*. Springer-Verlag, 2003.

Gordley, James. *The jurists: a critical history*. OUP Oxford, 2013.

Graciarena, Martin, Elizabeth Shriberg, Andreas Stolcke, Frank Enos, Julia Hirschberg, and Sachin Kajarekar. "Combining prosodic lexical and cepstral systems for deceptive speech detection." In 2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings, vol. 1, pp. I-I. IEEE, 2006.

Greene, Derek, Derek O'Callaghan, and Pádraig Cunningham. "How many topics? stability analysis for topic models." In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2014, Nancy, France, September 15-19, 2014. Proceedings, Part I 14*, pp. 498-513. Springer Berlin Heidelberg, 2014.

Gregory Jr, Stanford W., and Stephen Webster. "A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perceptions." *Journal of personality and social psychology* 70, no. 6 (1996): 1231.

Guyon, Isabelle, Jason Weston, Stephen Barnhill, and Vladimir Vapnik. "Gene selection for cancer classification using support vector machines." Machine learning 46 (2002): 389-422.

Hahn, Michael, and Navin Goyal. "A theory of emergent in-context learning as implicit structure induction." *arXiv preprint arXiv:2303.07971* (2023).

Haidt, Jonathan. *The righteous mind: Why good people are divided by politics and religion*. Vintage, 2012.

Hales, Steven D., and Rex Welshon. *Nietzsche's perspectivism*. University of Illinois Press, 2000.

Hamilton, William L., Jure Leskovec, and Dan Jurafsky. "Diachronic word embeddings reveal statistical laws of semantic change." *arXiv preprint arXiv:1605.09096* (2016).

Hammond, Kenneth R. "Coherence and correspondence theories in judgment and decision making." (2000).

Hanretty, Chris. A court of specialists: Judicial behavior on the UK Supreme Court. Oxford University Press, USA, 2020.

Hartmann, Jochen, Jasper Schwenzow, and Maximilian Witte. "The political ideology of conversational AI: Converging evidence on ChatGPT's pro-environmental, left-libertarian orientation." *arXiv preprint arXiv:2301.01768* (2023).

He, Haibo, and Edwardo A. Garcia. "Learning from imbalanced data." IEEE Transactions on knowledge and data engineering 21, no. 9 (2009): 1263-1284.

He, Haibo, Yang Bai, Edwardo A. Garcia, and Shutao Li. "ADASYN: Adaptive synthetic sampling approach for imbalanced learning." In 2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence), pp. 1322-1328. Ieee, 2008.

He, Haibo, and Yunqian Ma, eds. "Imbalanced learning: foundations, algorithms, and applications." (2013).

Hilliard, Airlie, Nigel Guenole, and Franziska Leutner. "Robots are judging me: Perceived fairness of algorithmic recruitment tools." *Frontiers in Psychology* 13 (2022): 940456.

Hoeflich, Michael H. "Law & geometry: Legal science from Leibniz to Langdell." *American Journal of Legal History* 30, no. 2 (1986): 95-121.

Holtzman, Ari, Jan Buys, Li Du, Maxwell Forbes, and Yejin Choi. "The curious case of neural text degeneration." *arXiv preprint arXiv:1904.09751* (2019).

Huang, Allen H., Hui Wang, and Yi Yang. "FinBERT: A large language model for extracting information from financial text." *Contemporary Accounting Research* 40, no. 2 (2023): 806-841.

Hume, David. *A treatise of human nature*. Clarendon Press, 1896.

Jaworska, Sylvia, and Anupam Nanda. "Doing well by talking good: A topic modelling-assisted discourse study of corporate social responsibility." *Applied Linguistics* 39, no. 3 (2018): 373-399.

Jaworska, Sylvia. "Change but no climate change: Discourses of climate change in corporate social responsibility reporting in the oil industry." *International Journal of Business Communication* 55, no. 2 (2018): 194-219.

Jaworska, Sylvia. "Corporate discourse." (2020).

Ji, Ziwei, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. "Survey of hallucination in natural language generation." *ACM Computing Surveys* 55, no. 12 (2023): 1-38.

Jiang, Zhiying, Matthew YR Yang, Mikhail Tsirlin, Raphael Tang, and Jimmy Lin. "Less is More: Parameter-Free Text Classification with Gzip." *arXiv preprint arXiv:2212.09410* (2022).

Joachims, Thorsten. "Text categorization with support vector machines: Learning with many relevant features." European conference on machine learning. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998.

Jones, Steven E. *Against technology: From the Luddites to neo-Luddism*. Routledge, 2013.

Jurafsky, Daniel, and James H. Martin. "Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition." 2023 ed.

Karttunen, Lauri, Jean-Pierre Chanod, Gregory Grefenstette, and Anne Schille. "Regular expressions for language engineering." Natural Language Engineering 2, no. 4 (1996): 305-328.

Kastellec, Jonathan P., Jeffrey R. Lax, Michael Malecki, and Justin H. Phillips. "Polarizing the electoral connection: partisan representation in Supreme Court confirmation politics." The journal of politics 77, no. 3 (2015): 787-804.

Kelsen, Hans. *Pure theory of law*. Univ of California Press, 1967.

Keynes, John Maynard. *They Economic Possibilities of Our Grandchildren* (1930)

Khalilia, Mohammed, Sounak Chakraborty, and Mihail Popescu. "Predicting disease risks from highly imbalanced data using random forest." BMC medical informatics and decision making 11 (2011): 1-13.

Kim, Hyogon, Eunmi Lee, and Donghee Yoo. "Do SEC filings indicate any trends? Evidence from the sentiment distribution of forms 10-K and 10-Q with FinBERT." *Data Technologies and Applications* ahead-of-print (2023).

Kirk, Hannah Rose, Yennie Jun, Filippo Volpin, Haider Iqbal, Elias Benussi, Frederic Dreyer, Aleksandar Shtedritski, and Yuki Asano. "Bias out-of-the-box: An empirical analysis of intersectional occupational biases in popular generative language models." *Advances in neural information processing systems* 34 (2021): 2611-2624.

Kozlowski, Austin C., Matt Taddy, and James A. Evans. "The geometry of culture: Analyzing the meanings of class through word embeddings." *American Sociological Review* 84, no. 5 (2019): 905-949.

Knoepfel, Ivo. "Who Cares Wins: Connecting Financial Markets to a Changing World, UN Environment Programme." (2004).

Kojima, Takeshi, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. "Large language models are zero-shot reasoners." *Advances in neural information processing systems* 35 (2022): 22199-22213.

Kruglanski, Arie W., and Icek Ajzen. "Bias and error in human judgment." *European Journal of Social Psychology* 13, no. 1 (1983): 1-44.

Kwak, Haewoon, Jisun An, Elise Jing, and Yong-Yeol Ahn. "FrameAxis: characterizing microframe bias and intensity with word embedding." *PeerJ Computer Science* 7 (2021): e644.

Lain, Corinna Barrett. "Deciding death." Duke LJ 57 (2007): 1. p 75

Laine, Matias. "Meanings of the term 'sustainable development' in Finnish corporate disclosures." In *Accounting Forum*, vol. 29, no. 4, pp. 395-413.

Landi, Giovanni, and Mauro Sciarelli. "Towards a more ethical market: the impact of ESG rating on corporate financial performance." *Social responsibility journal* 15, no. 1 (2018): 11-27.

Lazaridou, Angeliki, Adhi Kuncoro, Elena Gribovskaya, Devang Agrawal, Adam Liska, Tayfun Terzi, Mai Gimenez et al. "Mind the gap: Assessing temporal generalization in neural language models." *Advances in Neural Information Processing Systems* 34 (2021): 29348-29363.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." *nature* 521, no. 7553 (2015): 436-444.

Lee, Daniel D., and H. Sebastian Seung. "Learning the parts of objects by non-negative matrix factorization." Nature 401, no. 6755 (1999): 788-791.

Leibniz, Gottfried Wilhelm. "Elements of Natural Law." In *Philosophical papers and letters*, pp. 131-138. Springer Netherlands, 1989. 2nd edition Translated by: Leroy E. Loemker

Leiman, Joan Maisel. "The Rule of Four." Colum. L. Rev. 57 (1957): 975.

Lewis, Craig, and Steven Young. "Fad or future? Automated analysis of financial text and its implications for corporate reporting." *Accounting and Business Research* 49, no. 5 (2019): 587-615.

Li, Feng. "Textual analysis of corporate disclosures: A survey of the literature." *Journal of accounting literature* 29 (2010): 143.

Li, Jiazheng, Linyi Yang, Barry Smyth, and Ruihai Dong. "Maec: A multimodal aligned earnings conference call dataset for financial risk prediction." In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pp. 3063-3070. 2020.

Li, Ting-Ting, Kai Wang, Toshiyuki Sueyoshi, and Derek D. Wang. "ESG: Research progress and future prospects." *Sustainability* 13, no. 21 (2021): 11663.

Lilleberg, Joseph, Yun Zhu, and Yanqing Zhang. "Support vector machines and word2vec for text classification with semantic features." In 2015 IEEE 14th International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC), pp. 136-140. IEEE, 2015.

Lin, Yuxiao, Yuxian Meng, Xiaofei Sun, Qinghong Han, Kun Kuang, Jiwei Li, and Fei Wu. "Bertgcn: Transductive text classification by combining gcn and bert." arXiv preprint arXiv:2105.05727 (2021).

Linzer, Peter. "The meaning of certiorari denials." Colum L. Rev. 79 (1979): 1227.

López, Victoria, Alberto Fernández, and Francisco Herrera. "On the importance of the validation technique for classification with imbalanced datasets: Addressing covariate shift when data is skewed." Information Sciences 257 (2014): 1-13.

López, Victoria, Alberto Fernández, Salvador García, Vasile Palade, and Francisco Herrera. "An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics." Information sciences 250 (2013): 113-141.

Lu, Kaiji, Piotr Mardziel, Fangjing Wu, Preetam Amancharla, and Anupam Datta. "Gender bias in neural natural language processing." *Logic, Language, and Security: Essays Dedicated to Andre Scedrov on the Occasion of His 65th Birthday* (2020): 189-202.

Luhn, Hans Peter. "A statistical approach to mechanized encoding and searching of literary information." IBM Journal of research and development 1, no. 4 (1957): 309-317.

Luitse, Dieuwertje, and Wiebke Denkena. "The great transformer: Examining the role of large language models in the political economy of AI." *Big Data & Society* 8, no. 2 (2021): 20539517211047734.

Lund, Dorothy S., and Elizabeth Pollman. "The corporate governance machine." *Colum. L. Rev.* 121 (2021): 2563.

Madanagopal, Karthic, and James Caverlee. "Reinforced Sequence Training based Subjective Bias Correction." In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 2577-2590. 2023.

Mantyla, Mika V., Maelick Claes, and Umar Farooq. "Measuring LDA topic stability from clusters of replicated runs." In *Proceedings of the 12th ACM/IEEE international symposium on empirical software engineering and measurement*, pp. 1-4. 2018.

Martin, Andrew D., and Kevin M. Quinn. "Dynamic ideal point estimation via Markov chain Monte Carlo for the US Supreme Court, 1953–1999." Political analysis 10, no. 2 (2002): 134-153.

Maynez, Joshua, Shashi Narayan, Bernd Bohnet, and Ryan McDonald. "On faithfulness and factuality in abstractive summarization." *arXiv preprint arXiv:2005.00661* (2020).

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. "Distributed representations of words and phrases and their compositionality." *Advances in neural information processing systems* 26 (2013).

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. "Efficient estimation of word representations in vector space." *arXiv preprint arXiv:1301.3781* (2013).

Miller, George A. "WordNet: a lexical database for English." Communications of the ACM 38, no. 11 (1995): 39-41.

Monroe, Burt L., Michael P. Colaresi, and Kevin M. Quinn. "Fightin' words: Lexical feature selection and evaluation for identifying the content of political conflict." Political Analysis 16, no. 4 (2008): 372-403;

Morison, John, and Adam Harkens. "Re-engineering justice? Robot judges, computerised courts and (semi) automated legal decision-making." *Legal Studies* 39, no. 4 (2019): 618-635.

Mulligan, Thomas. "A critique of Milton Friedman's essay ‘the social responsibility of business is to increase its profits’." *Journal of Business Ethics* 5 (1986): 265-269.

Nietzsche, Freidrich, *Will to Power* (1901/1968), Vintage, translated by Kaufman

Nietzsche, Freidrich, *The Gay Science* (1882) trans. Josefine Nauckhoff, Cambridge University Press, (2001)

O'callaghan, Derek, Derek Greene, Joe Carthy, and Pádraig Cunningham. "An analysis of the coherence of descriptors in topic modeling." Expert Systems with Applications 42, no. 13 (2015): 5645-5657.

Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. "Thumbs up? Sentiment classification using machine learning techniques." arXiv preprint cs/0205070 (2002).

Parikh, Ankur P., Xuezhi Wang, Sebastian Gehrmann, Manaal Faruqui, Bhuwan Dhingra, Diyi Yang, and Dipanjan Das. "ToTTo: A controlled table-to-text generation dataset." *arXiv preprint arXiv:2004.14373* (2020).

Parikh, Purvish M., Dinesh M. Shah, and Kairav P. Parikh. "Judge Juan Manuel Padilla Garcia, ChatGPT, and a controversial medicolegal milestone." *Indian Journal of Medical Sciences* 75, no. 1 (2023): 3-8.

Pascal, Blaise. "Pensées", edited and translated by Roger Ariew (Hackett, 2004)

Peirce, Charles Santiago Sanders. "Prolegomena to an apology for pragmaticism." The Monist (1906): 492-546.

Peng, Baolin, Michel Galley, Pengcheng He, Hao Cheng, Yujia Xie, Yu Hu, Qiuyuan Huang et al. "Check your facts and try again: Improving large language models with external knowledge and automated feedback." *arXiv preprint arXiv:2302.12813* (2023).

Pérez, Lucy, Vivian Hunt, Hamid Samandari, Robin Nuttall, and Krysta Biniek. "Does ESG really matter—and why." *McKinsey Quarterly* (2022).

Perry, Hersel W. Deciding to decide: agenda setting in the United States Supreme Court. Harvard University Press, 2009.

Picard, Richard R., and R. Dennis Cook. "Cross-Validation of Regression Models." Journal of the American Statistical Association 79, no. 387 (1984): 575–83. https://doi.org/10.2307/2288403.

Pires, Telmo, Eva Schlinger, and Dan Garrette. "How multilingual is multilingual BERT?." *arXiv preprint arXiv:1906.01502* (2019).

Polonioli, Andrea. "The uses and abuses of the coherence–correspondence distinction." *Frontiers in Psychology* 6 (2015): 507.

Portugal, Ivens, Paulo Alencar, and Donald Cowan. "The use of machine learning algorithms in recommender systems: A systematic review." *Expert Systems with Applications* 97 (2018): 205-227.

Propp, Vladimir. *Morphology of the Folktale*. University of texas Press, 1968.

Propp, Vladimir. *Theory and history of folklore*. Vol. 5. U of Minnesota Press, 1985.

Qin, Yu, and Yi Yang. "What you say and how you say it matters: Predicting stock volatility using verbal and vocal cues." In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 390-401. 2019.

Quinn, Kevin M., Burt L. Monroe, Michael Colaresi, Michael H. Crespin, and Dragomir R. Radev. "How to analyze political attention with minimal assumptions and costs." American Journal of Political Science 54, no. 1 (2010): 209-228.

Quinonero-Candela, Joaquin, Masashi Sugiyama, Anton Schwaighofer, and Neil D. Lawrence, eds. Dataset shift in machine learning. Mit Press, 2008.

Radford, Alec, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. "Improving language understanding by generative pre-training." (2018).

Raghunandan, Aneesh, and Shivaram Rajgopal. "Do socially responsible firms walk the talk?." *Available at SSRN 3609056* (2022).

Ramezan, A, Christopher, Timothy A. Warner, and Aaron E. Maxwell. "Evaluation of sampling and cross-validation tuning strategies for regional-scale machine learning classification." Remote Sensing 11, no. 2 (2019): 185.

Recasens, Marta, Cristian Danescu-Niculescu-Mizil, and Dan Jurafsky. "Linguistic models for analyzing and detecting biased language." In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1650-1659. 2013.

Reinert, Hugo, and Erik S. Reinert. "Creative destruction in economics: Nietzsche, Sombart, schumpeter." *Friedrich Nietzsche (1844–1900) Economy and Society* (2006): 55-85.

Ren, Xiaozhe, Pingyi Zhou, Xinfan Meng, Xinjing Huang, Yadao Wang, Weichao Wang, Pengfei Li et al. "PanGu-{\Sigma}: Towards Trillion Parameter Language Model with Sparse Heterogeneous Computing." *arXiv preprint arXiv:2303.10845* (2023).

Revesz, Richard L., and Pamela S. Karlan. "Nonmajority Rules and the Supreme Court." U. Pa. L. Rev. 136 (1987)

Robbins, Ira P. "Justice by the Numbers: The Supreme Court and the Rule of Four-Or Is It Five." Suffolk UL Rev. 36 (2002)

Rock, Edward B. "For Whom is the Corporation Managed in 2020?: The Debate over Corporate Purpose." European Corporate Governance Institute-Law Working Paper 515 (2020): 20-16.

Rogers Anna, Olga Kovaleva, and Anna Rumshisky. "A primer in BERTology: What we know about how BERT works." *Transactions of the Association for Computational Linguistics* 8 (2021): 842-866.

Rozado, David. "The political biases of chatgpt." *Social Sciences* 12, no. 3 (2023): 148.

Ruder, Sebastian, Ivan Vulić, and Anders Søgaard. "A survey of cross-lingual word embedding models." *Journal of Artificial Intelligence Research* 65 (2019): 569-631.

Rutinowski, Jérôme, Sven Franke, Jan Endendyk, Ina Dormuth, and Markus Pauly. "The Self-Perception and Political Biases of ChatGPT." *arXiv preprint arXiv:2304.07333* (2023).

Saito, Takaya, and Marc Rehmsmeier. "The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets." *PloS one* 10, no. 3 (2015): e0118432.

Salton, Gerard, and Christopher Buckley. "Term-weighting approaches in automatic text retrieval." Information processing & management 24, no. 5 (1988): 513-523.

Samuel, Geoffrey. "Can legal reasoning be demystified?." *Legal Studies* 29, no. 2 (2009): 181-210.

Samuel, Geoffrey. "Is law really a social science? A view from comparative law." *The Cambridge Law Journal* 67, no. 2 (2008): 288-321.

Samuel, Geoffrey. "Is Legal Knowledge Regressing (Thanks to AI)?." *Amicus Curiae* 4, no. 3 (2023): 719-750.

Samuel, Geoffrey. "Taking Methods Seriously (Part One)." *J. Comp. L.* 2 (2007): 94.

Samuel, Geoffrey. "The challenge of artificial intelligence: can Roman law help us discover whether law is a system of rules?." *Legal Studies* 11, no. 1 (1991): 24-46.

Samuel, Geoffrey. *Epistemology and method in law*. Routledge, 2016.

Samuel, Geoffrey. *Rethinking legal reasoning*. Edward Elgar Publishing, 2018.

Santurkar, Shibani, Esin Durmus, Faisal Ladhak, Cinoo Lee, Percy Liang, and Tatsunori Hashimoto. "Whose opinions do language models reflect?." *arXiv preprint arXiv:2303.17548* (2023).

Schlechtweg, Dominik, Anna Hätty, Marco Del Tredici, and Sabine Schulte im Walde. "A wind of change: Detecting and evaluating lexical semantic change across times and domains." *arXiv preprint arXiv:1906.02979* (2019).

Schopenhauer, Arthur*: 'The World as Will and Representation': Volume 1* (1818). Cambridge University Press, 2010 ed. Trans: Judith Norman, Alistair Welchman and Christopher Janaway

Schwab, Klaus. *Stakeholder capitalism: A global economy that works for progress, people and planet*. John Wiley & Sons, 2021.

Sebastiani, Fabrizio. "Machine learning in automated text categorization." ACM computing surveys (CSUR) 34.1 (2002): 1-47.

Seneca, Lucius Annaeus, *Seneca: Moral and political essays*. Cambridge University Press, 1995.

Shan, Guogen. "Monte Carlo cross-validation for a study with binary outcome and limited sample size." BMC Medical Informatics and Decision Making 22, no. 1 (2022): 1-15.

Shao, Jun. "Linear model selection by cross-validation." Journal of the American statistical Association 88, no. 422 (1993): 486-494.

Sheng, Emily, Kai-Wei Chang, Premkumar Natarajan, and Nanyun Peng. "The woman worked as a babysitter: On biases in language generation." *arXiv preprint arXiv:1909.01326* (2019).

Sim, Yanchuan, Bryan Routledge, and Noah Smith. "The utility of text: the case of Amicus briefs and the Supreme Court." In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 29, no. 1. 2015.

Sourdin, Tania. "Judge v Robot?: Artificial intelligence and judicial decision-making." *University of New South Wales Law Journal, The* 41, no. 4 (2018): 1114-1133.

Sprinkle, Geoffrey B., and Laureen A. Maines. "The benefits and costs of corporate social responsibility." *Business Horizons* 53, no. 5 (2010): 445.

Stechemesser, Kristin, and Edeltraud Guenther. "Carbon accounting: a systematic literature review." *Journal of Cleaner Production* 36 (2012): 17-38.

Stephenson, Ryan. "Federal Circuit Cases Selection at the Supreme Court: An Empirical Analysis." Geo. LJ 102 (2013): 271.

Stern, Robert L., and Eugene Gressman. Supreme Court practice: jurisdiction, procedure, arguing and briefing techniques, forms, statutes, rules for practice in the Supreme Court of the United States. Bureau of national affairs, 1985.

Strine Jr, Leo E. "The dangers of denial: The need for a clear-eyed understanding of the power and accountability structure established by the Delaware general corporation law." *Wake Forest L. Rev.* 50 (2015): 761.

Strine Jr, Leo E., Justin L Brooke, Kyle M. Diamond, and Derrick L. Parker Jr. "It's Time to Focus on the “G” in ESG." *Harvard Business Review, November* 18 (2022).

Strubell, Emma, Ananya Ganesh, and Andrew McCallum. "Energy and policy considerations for deep learning in NLP." *arXiv preprint arXiv:1906.02243* (2019).

Suchman, Mark C. "Managing legitimacy: Strategic and institutional approaches." Academy of management review 20, no. 3 (1995): 571-610.

Sulea, Octavia-Maria, Marcos Zampieri, Mihaela Vela, and Josef Van Genabith. "Predicting the law area and decisions of French Supreme Court cases." arXiv preprint arXiv:1708.01681 (2017).

Susskind, Richard. *Online courts and the future of justice.* (2019).

Susskind, Richard. *Expert systems in law*. Oxford University Press, Inc., 1987.

Susskind, Richard. *Tomorrow's lawyers: An introduction to your future*. Oxford University Press, 2013.

Swets, John A. "Measuring the accuracy of diagnostic systems." *Science* 240, no. 4857 (1988): 1285-1293.

Tanenhaus, Joseph, et al. "The Supreme Court's certiorari jurisdiction: Cue theory." Judicial decision making 111 (1963).

Teger, Stuart H., and Douglas Kosinski. "The cue theory of Supreme Court certiorari jurisdiction: A reconsideration." The Journal of Politics 42.3 (1980): 834-846.

Turing, Alan M. *Computing machinery and intelligence*. Springer Netherlands, 1950/2009.

Ulmer, S. Sidney, William Hintze, and Louise Kirklosky. "The decision to grant or deny certiorari: Further consideration of cue theory." Law & Society Review 6.4 (1972): 637-643.

Ulmer, S. Sidney. "The decision to grant certiorari as an indicator to decision" on the merits"." Polity 4.4 (1972)

Vamathevan, Jessica, Dominic Clark, Paul Czodrowski, Ian Dunham, Edgardo Ferran, George Lee, Bin Li et al. "Applications of machine learning in drug discovery and development." *Nature reviews Drug discovery* 18, no. 6 (2019): 463-477.

van de Bijl, Etienne, Jan Klein, Joris Pries, Sandjai Bhulai, Mark Hoogendoorn, and Rob van der Mei. "The Dutch Draw: constructing a universal baseline for binary prediction models." *arXiv preprint arXiv:2203.13084* (2022).

Vapnik, Vladimir. The nature of statistical learning theory. Springer science & business media, 1999.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. "Attention is all you need." *Advances in neural information processing systems* 30 (2017).

Voigt, Rob, Nicholas P. Camp, Vinodkumar Prabhakaran, William L. Hamilton, Rebecca C. Hetey, Camilla M. Griffiths, David Jurgens, Dan Jurafsky, and Jennifer L. Eberhardt. "Language from police body camera footage shows racial disparities in officer respect." *Proceedings of the National Academy of Sciences* 114, no. 25 (2017): 6521-6526.

Volokh, Eugene. "Chief justice robots." *Duke LJ* 68 (2018): 1135.

von Savigny, Friedrich Carl. *Of the Vocation of Our Age for Legislation and Jurisprudence*. Littlewood and Company, 1831. Translated by: Abraham Hayward, p 38-39

Wang, Nu. "“Black Box Justice”: Robot Judges and AI-based Judgment Processes in China’s Court System." In *2020 IEEE International Symposium on Technology and Society (ISTAS)*, pp. 58-65. IEEE, 2020.

Watson, Alan trans ed. The Digest of Justinian. University of Pennsylvania Press, 2011.

Wei, Jason, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V. Le, and Denny Zhou. "Chain-of-thought prompting elicits reasoning in large language models." *Advances in Neural Information Processing Systems* 35 (2022): 24824-24837.

Wei, Jason, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama et al. "Emergent abilities of large language models." *arXiv preprint arXiv:2206.07682* (2022).

Weinstein, James. "The corporate ideal in the liberal state: 1900-1918." (1969).

Welleck, Sean, Ilia Kulikov, Stephen Roller, Emily Dinan, Kyunghyun Cho, and Jason Weston. "Neural text generation with unlikelihood training." *arXiv preprint arXiv:1908.04319* (2019).

Whelan, Tenise, U. Atz, T. Van Holt, and C. Clark. "Uncovering the Relationship by Aggregating Evidence from 1,000 Plus Studies Published between 2015–2020." 2022.

Widyawati, Luluk. "A systematic literature review of socially responsible investment and environmental social governance metrics." *Business Strategy and the Environment* 29, no. 2 (2020): 619-637.

Xia, Yuqiang, and Michel C. Boufadel. "Lessons from the Exxon Valdez oil spill disaster in Alaska." *Disaster Advances* 3, no. 4 (2010): 270-273.

Xu, Qing-Song, and Yi-Zeng Liang. "Monte Carlo cross validation." Chemometrics and Intelligent Laboratory Systems 56, no. 1 (2001): 1-11.

Xu, Zichun. "Human Judges in the era of artificial intelligence: challenges and opportunities." *Applied Artificial Intelligence* 36, no. 1 (2022): 2013652.
Yu, Ellen Pei‐yi, Christine Qian Guo, and Bac Van Luu. "Environmental, social and governance transparency and firm value." *Business Strategy and the Environment* 27, no. 7 (2018): 987-1004.

Zellers, Rowan, Ari Holtzman, Hannah Rashkin, Yonatan Bisk, Ali Farhadi, Franziska Roesner, and Yejin Choi. "Defending against neural fake news." *Advances in neural information processing systems* 32 (2019).

Zhao, Wayne Xin, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min et al. "A survey of large language models." *arXiv preprint arXiv:2303.18223* (2023).

Zhao, Weizhong, James J. Chen, Roger Perkins, Zhichao Liu, Weigong Ge, Yijun Ding, and Wen Zou. "A heuristic approach to determine an appropriate number of topics in topic modeling." In *BMC bioinformatics*, vol. 16, pp. 1-10. BioMed Central, 2015.

Zhou, Chunting, Graham Neubig, Jiatao Gu, Mona Diab, Paco Guzman, Luke Zettlemoyer, and Marjan Ghazvininejad. "Detecting hallucinated content in conditional neural sequence generation." *arXiv preprint arXiv:2011.02593* (2020).

Žižek, Slavoj. *Event: Philosophy in transit*. Penguin UK, 2014.

Žižek, Slavoj. *First as tragedy, then as farce*. Verso, 2009.

Zizek, Slavoj. *The sublime object of ideology*. Verso Books, 2009