

# Can We Algorithmize Politics? The Promise and Perils of Computerized Text Analysis in Political Research

Mor Mitrani, *Bar Ilan University, Israel*

Tracy Adams, *Yale University, USA*

Inbar Noy, *The Hebrew University of Jerusalem, Israel*

## ABSTRACT

In recent years, political scientists increasingly have used data-science tools to research political processes, positions, and behaviors. Because both domestic and international politics are grounded in oral and written texts, computerized text analysis (CTA)—typically based on natural-language processing—has become one of the most notable applications of data-science tools in political research. This article explores the promises and perils of using CTA methods in political research and, specifically, the study of international relations. We highlight fundamental analytical and methodological gaps that hinder application and review processes. Whereas we acknowledge the significant contribution of CTA to political research, we identify a dual “engagement deficit” that may distance those without prior background in data science: (1) the tendency to prioritize methodological innovation over analytical and theoretical insights; and (2) the scholarly and political costs of requiring high proficiency levels and training to comprehend, assess, and use advanced research models.

Scientific progress often is contingent on methodological innovation. Unlike theories and empirical data, methods typically are more prone to migration across disciplines because they are, by nature, more adaptive and less associated with a concrete scholarly field. However, importing methods from other scholarly fields rarely is self-sufficient; method migration often involves theoretical and analytical modifications that design and reshape research programs, and it is highly contingent on the host discipline.

As part of the significant turn to computational social sciences in the past two decades, we have witnessed a growing scholarship

that adopts data-science tools in political research and interweaves cutting-edge computational perspectives with substantial questions on political processes, positions, and behavior. Given the extensive role of both oral and written texts and interactions in political doing and making, natural language processing (NLP)–based methods of computerized text analysis (CTA) have gained notable prominence, mainly in the fields of comparative politics, American politics, and electoral studies (Schuler 2020; Wilkerson and Casas 2017). In international relations (IR), however, the use of these methods is still somewhat nascent.

IR is a relatively young field of research that from its inception was—and still is—heavily influenced by other disciplines, both theoretically and methodologically (Schmidt 2016). IR often is slow to respond to trends that dominate other branches in the broader scholarship of political research. Thus, although there has been a growing interest in using computerized methods to analyze international data in recent years, applying these tools to examine IR research objectives has not

**Mor Mitrani**  is a lecturer of political studies at Bar Ilan University. She can be reached at [mor.mitrani@biu.ac.il](mailto:mor.mitrani@biu.ac.il).

**Tracy Adams**  is a postdoctoral researcher in the Center for Cultural Sociology at Yale University. She can be reached at [tracy.adams@yale.edu](mailto:tracy.adams@yale.edu).

**Inbar Noy**  is a PhD candidate in international relations at The Hebrew University of Jerusalem. She can be reached at [inbar.noy@mail.huji.ac.il](mailto:inbar.noy@mail.huji.ac.il).

© The Author(s), 2022. Published by Cambridge University Press on behalf of the American Political Science Association. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

doi:10.1017/S1049096522000464

PS • October 2022 809

yet met its full potential. As this article demonstrates, the case of applying CTA to the IR field allows us to closely examine the migration of methods from one field to another and to assess the accompanying possibilities and hurdles. The main challenge, we argue, is not the introduction of these new methods, which can be measured simply by the extent to which scholars adopt CTA methods in their research, but rather—and more important—their precarious engagement and application.

This article questions the usually positive perspective on the ability of computational methods to boost research in social sciences at large and political science and IR in particular in terms of volume, variety, velocity, and vinculation, thereby promoting innovation in data-collection data analysis (Monroe 2013). We fully acknowledge that political science, like many other disciplines, is on the cusp of a transition to an academic world in which artificial intelligence (AI) knowledge and machine-learning methodologies are an integral part of research programs. However, we demonstrate that computational models often are borrowed and methodologically implemented without giving due attention to the analytical context. The insufficient tailoring of these methods to the “receiving” field often results in studies that rely heavily on code and thus are approachable and transparent only to those few scholars who master computer language. Therefore, despite the promise of computational methods, we caution against their unquestioning application. We highlight two main caveats regard-

*...computational models often are borrowed and methodologically implemented without giving due attention to the analytical context.*

ing the import of computational-method packages without careful adaptations: (1) the prioritization of methodological innovation at the expense of analytical substance; and (2) a growing inaccessibility and lack of transparency. We discuss possible options for mitigating and overcoming potential discrepancies and complexities, highlighting the responsibility of the scholarly community to consider both the analytical challenge of the computational turn and its potential political ramifications—namely, widening existing gaps and creating digital inequality.

#### BRINGING CTA TO POLITICAL RESEARCH: THE CASE OF IR

The eminent spread of digital interactions, social networks, and online activities that have reshaped our social habitat is encouraging researchers across disciplines to rethink and revise the main paradigmatic frameworks of social and political research (Jungherr and Theocharis 2017, 99; Lazer et al. 2009, 722. Indeed, in recent years, political scientists have used digital datafication trends (Mayer-Schonberger and Cukier 2014) to introduce new types of data and compile an incredible array of new databases (Grossman and Pedahzur 2020, 226). Computational social sciences harness the use and spread of big data and machine-learning tools for modeling, simulating, and scrutinizing social phenomena by computational means (Brady 2019, 297–98). They enable the analysis of high-dimensional and noisy datasets and provide new insights into thus far latent and unreachable layers of social and political life (González-Bailón 2013, 153). Political scientists also have implemented and developed computational models based on AI and machine learning for exploring various political

phenomena (see Chatsiou and Mikhaylov 2020 for an excellent review): for example, a forecast model for predicting US election results (Linzer 2013) and an estimation model of candidates' ideologies and levels of endorsement (Bond and Messing 2015).

One of the most notable contributions of the import of data science to the political field is the introduction and development of the “text-as-data” approach to political science (Grimmer and Stewart 2013). This approach acknowledges the promise of advanced tools for automatically collecting substantial amounts of texts and analyzing the patterns of talk and speech that characterize and constitute political realms. Political scientists have used CTA to analyze a wide range of political corpora, including party manifestos (e.g., Benoit et al. 2016; Benoit, Laver, and Mikhaylov 2009; Dinas and Gemenis 2010) and speeches (e.g., Beata, Diermeier, and Beigman 2008; Lauderdale and Herzog 2016; Wiener 2007), and to develop models for the automatic measuring, scoring, and scaling of political actors' positions and preferences, including parties, legislators, and interest groups (Grimmer 2010; Laver, Benoit, and Garry 2003; Roberts et al. 2014; Slapin and Proksch 2008).

In the IR field, the potential of CTA for text analysis is indisputable. The international political sphere is rich in texts and built of texts, relying on and realized by discursive and textual interactions. Public discourse at the international level is an essential source of data, and computerized methods can foster

systematic examination of the interactions that ultimately design our primary subject matter: world politics. Indeed, in recent years, we have witnessed nascent albeit burgeoning literature applying CTA-based research to various corpora: nongovernmental-organization reports (e.g., Fariss et al. 2015; Park, Murdie, and Davis 2019); international investment agreements (Alschner and Skougarevskiy 2016); international climate-change negotiations (Bagozzi 2015); the United Nations Security Council (Schönfeld et al. 2019); the United Nations General Debate (UNGD) corpus (see, e.g., Baturu, Dasandi, and Mikhaylov 2017; Chelotti, Dasandi, and Mikhaylov 2021; Dieng, Ruiz, and Blei 2019; Gurciullo and Mikhaylov 2017a; Watanabe and Zhou 2020); and academic discourse in IR journals (Steffek, Müller, and Behr 2021; Whyte 2019).

However, despite the increasing interest in CTA in IR, examination of the relevant research reveals a dual “engagement deficit.” First, the objective of most of these applications is, primarily, methodological and thus directed at developing data-science models rather than advancing existing knowledge and analytical purviews of IR. Second, and related, they rely heavily on a computational language that requires proficiency, thereby reducing the chances that non-data-science-trained scholars can fully comprehend.

#### INSIGHTS FROM THE UNITED NATIONS GENERAL DEBATE CORPUS

In recent years, much scholarly attention has been given to the previously neglected corpus of speeches in the annual general debate of the United Nations General Assembly. In international

---

politics, the UNGD is a rare and perhaps the only ritualistic discursive arena in which states have convened regularly and equally since 1945. Despite its name, it is less a debate and more a battery of speeches typically delivered by heads of state in a highly structured and ritualized way. These texts often signify states' perceptions and experiences of world affairs, thus serving as a barometer (Smith 2006) that traces the agenda of international politics (Mingst and Karns 2011). IR researchers tend to show little interest in these speeches. However, there has been recent systematic quantitative and qualitative research of this textual corpus (Baturu, Dasandi, and Mikhaylov 2017; Hecht 2016; Kentikelenis and Voeten 2021)<sup>1</sup> that highlights these texts as a promising data source for illuminating latent currents in world politics and teaches us about the dynamics of international discourse.

Not only IR scholars have found interest in this corpus; in recent years, data scientists also have presented and published several studies applying various NLP methods to this dataset. However, most of the studies were conducted by data scientists who published or presented them in data science journals, archives, and conferences (e.g., Blei and McAuliffe 2010; Dieng, Ruiz, and Blei 2019; Gurciullo and Mikhaylov 2017a, 2017b), thereby advancing computational development more than political knowledge. Blei's works are a notable example. A prominent computer scientist at Columbia University, Blei and his colleagues use political corpora (including the UNGD) to develop NLP algorithms for textual analysis. Their work is directed almost exclusively to the data-science community; therefore, their publications also remain in this realm. Even Watanabe and Zhou's (2020) attempt to directly address IR scholars by showing how semi-supervised methods may assist theory-driven analysis in IR eventually was published in *Social Science Computer Review*, a non-IR journal. Consequently, a political scientist who wants to build on these studies to advance political theories would have to invest significant effort to locate and much less understand them. Although efforts have been made to suggest potential political insights (e.g., Baturu, Dasandi, and Mikhaylov 2017; Chelotti, Dasandi, and Mikhaylov 2021), these studies primarily emphasized the technical elements of applying CTA methods and models.

The tension between analytical and methodological components of research is well known. Returning to the fundamentals of research, we know that an analytical framework is a prerequisite and that research questions should guide the decisions made about both the method and the analysis. For a researcher, however, the choice of method—especially in a world of big data and automated code-based analysis—is like being in a magical theme

“conflict of interests,” they may be detached from an analytical anchor and thus unable to deliver in terms of promoting analytical, theoretical, and empirical insights to the IR field.

The problem is intensified further when many of the methodological models used are not sufficiently sensitive to the domain that they are designed to analyze. In principle, organizing data for computational analysis requires attention to domain-specific issues and poses limitations in both the pre-processing and processing phases (Denny and Spirling 2018, 170). Analyzing international political texts, which are rich in the presence of unique entities such as the names of political leaders, states, nationalities, organizations, and legal texts, requires researchers to rely on more than standard tools. They must be fully acquainted with the political concepts and terms, and they subsequently must train the models to recognize them lest their findings be distorted and fail to represent the data accurately and validly. In our experience, many studies—despite lengthy methodological indices—lack much-needed transparency regarding the decisions made throughout the initial organizing, cleaning, and pre-processing of the texts.<sup>2</sup> Consequently, this limits their ability to assess the political nature of the texts.

#### PATHWAYS FORWARD: CAN WE SIMULTANEOUSLY BE A DATA SCIENTIST AND A POLITICAL SCIENTIST?

The “big-data revolution” is more than simply a trendy buzzword. It affects every aspect of society and, therefore, politics, and it provides promising opportunities for research across disciplines. Method and methodological innovation ultimately should complement one another and not replace the need for theoretical and analytical frameworks. This is not a novel idea. The potential challenges of applying CTA in particular and big-data analysis in general to political research were identified previously. It is well established that data alone cannot “speak for itself” and that political researchers are obliged to not only reshape traditional methods of data collection and analysis but also to “rethink how they do political science” (Brady 2019, 298), considering that theory always is needed to shed light on the complex political phenomena being examined (Grimmer 2014, 81–82; Kitchin 2014, 2; Titiunik 2015, 76). Although we mainly refer to examples from the IR discipline, they are nonetheless relevant and valid for political science at large and social sciences as a whole.

We join these cautioning voices and specifically illuminate the professional cost (and value) of introducing and relying on foreign programming languages. As demonstrated in this article, such analyses often are conducted by researchers who specialize in computerized methods but not necessarily political science; con-

*For a researcher, the choice of method—especially in a world of big data and automated code-based analysis—is like being in a magical theme park packed with inordinate possibilities.*

park packed with inordinate possibilities. The challenge is even greater when fields and disciplines collide; data scientists are prone to advancing ways to collect and utilize data of any type, whereas IR scholars are oriented toward gaining political insight and knowledge. Whereas many of the studies present sophisticated and cutting-edge methodologies, with this potential

sequently, many CTA applications prioritize methodology innovation over immersion in the political field. There is no doubt that methodological innovation is critical and essential for enriching the political-research toolkit. However, we should be aware that (1) this innovation may come at the expense of providing new empirical and theoretical insights; and (2) the ability of scholars

who are not trained in computerized methods to review, assess, and even understand the research process is limited. For example, in many CTA-based papers that are published in prominent political science journals (e.g., Barnum and Lo 2020; Greene and Cross 2017; Park, Greene, and Colaresi 2020), the design, execution, and language used often are rich in professional jargon, thereby possibly hindering and even preventing engagement with wide audiences within the political science community. This may quickly distance those (many) political researchers who have no prior knowledge of data science and may result not only in low-quality or even inaccurate research but also engender a publication bias that promotes proficiency in computer science over political science. Ultimately, importing ready-made method packages from external fields and disciplines as new methodological purviews for analyzing politics is an obstacle not only because it minimizes the potential reach of these methods but also because the solution cannot be limited to increased training.

In response to different methodological trends, political science graduate students have been trained in the past two decades in advanced statistics, experimental designs, and various software languages along with the core political science curriculum. At some point, developing these skills must come at the expense of deep and exhaustive knowledge of the dynamic political field and its research traditions. Moreover, not all political scientists have the privilege of learning and employing intricate text-as-data methods or have access to the costly hardware, software, and bandwidth that these methods demand. The challenge is not only the heavy burden of expanding the spectrum of training now required of political scientists; it also is—and perhaps even more—the invisible and thickening veil that separates those who can do the research and those who are supposed to understand and

common and balanced ground concerning the use of technical language and an analytical framework that can make studies in both disciplines more accessible. Review processes play a vital role in this; authors must be committed to a vocabulary that is comprehended easily and reviewers to a more hospitable approach toward new methods. This also requires transparency regarding the choices and decisions made throughout the research process (Kapiszewski and Karcher 2021)—for example, by providing explanations of the connection among the method, the results, and the political implications.

This also is pivotal for the application of NLP-based methods in IR; they continue to emerge and develop; thus, meticulous engagement approaches can be used to harness wider audiences within the IR community. However, these approaches require going beyond the promotion of inclusiveness and interdisciplinary collaborations. First and foremost, they require caution against conflating lack of knowledge with self-abnegation. New and unknown methods often are captivating but cannot and should not be followed blindly. Whereas data science—especially in “soft” political science—may appear to be a solution that provides objective and computerized tools that minimize human intervention and solve common issues of limited research choices, this ultimately is not the case. Eventually, computerized models and methods are constructed and decided by human intervention, and they are as subjective and biased as any other method (Chatsiou and Mikhaylov 2020). In fact, human interpretivism guided by political-oriented knowledge is a crucial part of developing more advanced and accurate computerized tools, particularly because—from the viewpoint of political scientists—texts and words cannot and should not be treated solely as a methodological resource for data. Text is a (some scholars would argue *the*) fundamental and primary polit-

*The challenge is not only the heavy burden of expanding the spectrum of training now required of political scientists; it also is—and perhaps even more—the invisible and thickening veil that separates those who can do the research and those who are supposed to understand and review it but are at a loss when it comes to deciphering long and cryptic Greek-letter formulas and code scripts.*

review it but are at a loss when it comes to deciphering long and cryptic Greek-letter formulas and code scripts.

This state of affairs has important political implications. Conducting and learning computational research are extremely costly and therefore available only to those few who are employed at or study in high-ranking, wealthy academic institutions that can provide access to the often-expensive program and facilities required for these endeavors. The more computational methods become a requisite for political research, the more this trend will widen scholarly inequalities by excluding groups of scholars who often already are underrepresented in major political science journals.

This article is not a call to resist evolution; advancing science relies on developing new research trajectories. Nonetheless, normalizing these questions and articulating skepticism can promote a more open, dialogic, and constructive research and highlight the need for interdisciplinary collaboration. The conditions for such a dialogue, first and foremost, depend on working together to find a

ical tool through which actors present identities, construct (political) relations, and do and make politics through various mechanisms such as legitimacy and identification. Thus, text is not only a methodological source for political research but also an epistemological construct through which actors understand, present, and conduct political relations (e.g., Carta 2019; Lundborg and Vaughan-Williams 2015). This is especially relevant in the international arena, which is less formal and hierarchical and therefore heavily shaped and reshaped through textual and discursive interactions among an excessive array of agents. Reducing political texts to serve only as variables and indicators narrows the potential scope of analysis and insight that text analysis can yield in political research in general and in IR in particular.

#### ACKNOWLEDGMENTS

For helpful comments and suggestions, the authors thank Jonathan Grossman, Mathis Lohaus, and the panel participants

and audience at the 2021 Virtual International Studies Association Conference, as well as the anonymous reviewers and *PS: Political Science & Politics* editors. This article is part of the research project, “What Are States Talking About?” (ISF Grant 2109/19), funded by the Israeli Science Foundation.

## CONFLICTS OF INTEREST

The authors declare that there are no ethical issues or conflicts of interest in this research. ■

## NOTES

1. In fact, Baturo, Dasandi, and Mikhaylov (2017) were the first to develop and introduce the code for mining the texts; until then, research conducted on the speeches required manual downloading and indexing.
2. For more on the importance of transparency in political science, see Jacobs, Kapiszewski, and Karcher (2022).

## REFERENCES

- Alschnner, Wolfgang, and Dmitriy Skougarevskiy. 2016. “Mapping the Universe of International Investment Agreements.” *Journal of International Economic Law* 19 (3): 561–88. <https://doi.org/10.1093/jiel/jgw056>.
- Bagozzi, Benjamin E. 2015. “The Multifaceted Nature of Global Climate-Change Negotiations.” *Review of International Organizations* 10 (4): 439–64. <https://doi.org/10.1007/s11558-014-9211-7>.
- Barnum, Miriam, and James Lo. 2020. “Is the NPT Unraveling? Evidence from Text Analysis of Review Conference Statements.” *Journal of Peace Research* 12 (1): 1–12.
- Baturo, Alexander, Niheer Dasandi, and Slava J. Mikhaylov. 2017. “Understanding State Preferences with Text as Data: Introducing the UN General Debate Corpus.” *Research and Politics* 4 (2): 1–7. <https://doi.org/10.1177/2053168017712821>.
- Beata, Beigman, Daniel Diermeier, and Eyal Beigman. 2008. “Lexical Cohesion Analysis of Political Speech.” *Political Analysis* 16 (4) (Special Issue): 447–63. <https://doi.org/10.1093/pan/mpn007>.
- Benoit, Kenneth, Drew Conway, Benjamin E. Lauderdale, Michael Laver, and Slava Mikhaylov. 2016. “Crowd-Sourced Text Analysis: Reproducible and Agile Production of Political Data.” *American Political Science Review* 110 (2): 278–95. <https://doi.org/10.1017/S0003055416000058>.
- Benoit, Kenneth, Mikhaylov Laver, and Slava Mikhaylov. 2009. “Treating Words as Data with Error: Uncertainty in Text Statements of Policy Positions.” *American Journal of Political Science* 53 (2): 495–513. <https://doi.org/10.1111/j.1540-5907.2009.00383.x>.
- Blei, David M., and Jon D. McAuliffe. 2010. “Supervised Topic Models.” *ArXiv Preprint ArXiv:1003.0783*, 1–22. <https://arxiv.org/abs/1003.0783>.
- Bond, Robert, and Solomon Messing. 2015. “Quantifying Social Media’s Political Space: Estimating Ideology from Publicly Revealed Preferences on Facebook.” *American Political Science Review* 109 (1): 62–78. <https://doi.org/10.1017/S0003055414000525>.
- Brady, Henry E. 2019. “The Challenge of Big Data and Data Science.” *Annual Review of Political Science* 22:297–323. <https://doi.org/10.1146/annurev-polisci-090216-023229>.
- Carta, Caterina. 2019. “‘A Rose by Any Other Name’: On Ways of Approaching Discourse Analysis.” *International Studies Review* 21 (1): 81–106.
- Chatsiou, Kakia, and Slava Jankin Mikhaylov. 2020. “Deep Learning for Political Science.” *ArXiv Preprint ArXiv:2005.06540*. <https://arxiv.org/abs/2005.06540>.
- Chelotti, Nicola, Niheer Dasandi, and Slava Jankin Mikhaylov. 2021. “Do Intergovernmental Organizations Have a Socialization Effect on Member State Preferences? Evidence from the UN General Debate.” *International Studies Quarterly* 66 (1): 1–17. <https://doi.org/10.1093/isq/sqab069>.
- Denny, Matthew J., and Arthur Spirling. 2018. “Text Pre-Processing for Unsupervised Learning: Why It Matters, When It Misleads, and What to Do About It.” *Political Analysis* 26 (2): 168–89. <https://doi.org/10.1017/pan.2017.44>.
- Dieng, Adji B., Francisco J. R. Ruiz, and David M. Blei. 2019. “The Dynamic Embedded Topic Model.” *ArXiv Preprint ArXiv:1907.05545*, 1–17. <https://arxiv.org/abs/1907.05545>.
- Dinas, Elias, and Kostas Gemenis. 2010. “Measuring Parties’ Ideological Positions with Manifesto Data: A Critical Evaluation of the Competing Methods.” *Party Politics* 16 (4): 427–50. <https://doi.org/10.1177/1354068809343107>.
- Fariss, Christopher J., Fridolin J. Linder, Zachary M. Jones, Charles D. Crabtree, Megan A. Biek, Ana-Sophia M. Ross, Taranamol Kaur, and Michael Tsai. 2015. “Human Rights Texts: Converting Human Rights Primary Source Documents into Data.” *PLOS One* 10 (9): e0138935. <https://doi.org/10.1371/journal.pone.0138935>.
- González-Bailón, Sandra. 2013. “Social Science in the Era of Big Data.” *Policy and Internet* 5 (2): 147–60. <https://doi.org/10.1002/1944-2866.POI328>.
- Greene, Derek, and James P. Cross. 2017. “Exploring the Political Agenda of the European Parliament Using a Dynamic Topic Modeling Approach.” *Political Analysis* 25 (1): 77–94.
- Grimmer, Justin. 2010. “A Bayesian Hierarchical Topic Model for Political Texts: Measuring Expressed Agendas in Senate Press Releases.” *Political Analysis* 18 (1): 1–35. <https://doi.org/10.1093/pan/mp034>.
- Grimmer, Justin. 2014. “We Are All Social Scientists Now: How Big Data, Machine Learning, and Causal Inference Work Together.” *PS: Political Science & Politics* 48 (1): 80–83. <https://doi.org/10.1017/S1049096514001784>.
- Grimmer, Justin, and Brandon M. Stewart. 2013. “Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts.” *Political Analysis* 21 (3): 267–97. <https://doi.org/10.1093/pan/mps028>.
- Grossman, Jonathan, and Ami Pedahzur. 2020. “Political Science and Big Data: Structured Data, Unstructured Data, and How to Use Them.” *Political Science Quarterly* 135 (2): 225–57. <https://doi.org/10.1002/polq.13032>.
- Gurciullo, Stefano, and Slava Mikhaylov. 2017a. “Topology Analysis of International Networks Based on Debates in the United Nations” *arXiv preprint*, 1–27. <https://arxiv.org/abs/1707.09491>.
- Gurciullo, Stefano, and Slava J. Mikhaylov. 2017b. “Detecting Policy Preferences and Dynamics in the UN General Debate with Neural Word Embeddings.” *International Conference on the Frontiers and Advances in Data Science*, 74–79. <https://arxiv.org/abs/1707.03490>.
- Hecht, Catherine. 2016. “The Shifting Salience of Democratic Governance: Evidence from the United Nations General Assembly General Debates.” *Review of International Studies* 42 (5): 915–38. <https://doi.org/10.1017/S0260210516000073>.
- Jacobs, Alan M., Diana Kapiszewski, and Sebastian Karcher. 2022. “Using Annotation for Transparent Inquiry (ATI) to Teach Qualitative Research Methods.” *PS: Political Science & Politics* 55 (1): 216–20. DOI:10.1017/S1049096521001335.
- Jungherr, Andreas, and Yannis Theocharis. 2017. “The Empiricist’s Challenge: Asking Meaningful Questions in Political Science in the Age of Big Data.” *Journal of Information Technology and Politics* 14 (2): 97–109. <https://doi.org/10.1080/19331681.2017.1312187>.
- Kapiszewski, Diana, and Sebastian Karcher. 2021. “Empowering Transparency: Annotation for Transparent Inquiry (ATI).” *PS: Political Science & Politics* 54 (3): 473–78. <https://doi.org/10.1017/S1049096521000287>.
- Kentikelenis, Alexander, and Erik Voeten. 2021. “Legitimacy Challenges to the Liberal World Order: Evidence from United Nations Speeches, 1970–2018.” *Review of International Organizations* 16 (4): 721–54.
- Kitchin, Rob. 2014. “Big Data, New Epistemologies, and Paradigm Shifts.” *Big Data and Society* 1 (1): 1–12. <https://doi.org/10.1177/2053951714528481>.
- Lauderdale, Benjamin E., and Alexander Herzog. 2016. “Measuring Political Positions from Legislative Speech.” *Political Analysis* 24 (3): 374–94. <https://doi.org/10.1093/pan/mpw017>.
- Laver, Michael, Kenneth Benoit, and John Garry. 2003. “Extracting Policy Positions from Political Texts Using Words as Data.” *American Political Science Review* 97 (2): 311–31. <https://doi.org/10.1017/S0003055403000698>.
- Lazer, David, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, and Nicholas Christakis. 2009. “Social Science: Computational Social Science.” *Science* 323 (5915): 721–23. DOI:10.1126/science.1167742.
- Linzer, Drew A. 2013. “Dynamic Bayesian Forecasting of Presidential Elections in the States.” *Journal of the American Statistical Association* 108 (501): 124–34. <https://doi.org/10.1080/01621459.2012.737735>.
- Lundborg, Tom, and Nick Vaughan-Williams. 2015. “New Materialisms, Discourse Analysis, and International Relations: A Radical Intertextual Approach.” *Review of International Studies* 41 (1): 3–25. <https://doi.org/10.1017/S0260210514000163>.
- Mayer-Schonberger, Viktor, and Kenneth Cukier. 2014. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Boston: Houghton Mifflin Harcourt.
- Mingst, Karen A., and Margaret P. Karns. 2011. *The United Nations in the 21st Century*. Boulder, CO: Westview Press.
- Monroe, Burt L. 2013. “The Five Vs of Big Data Political Science: Introduction to the Virtual Issue on Big Data in Political Science.” *Political Analysis* 21 (5): 1–9. <https://doi.org/10.1017/S1047198700014315>.

- Park, Baekkwon, Kevin Greene, and Michael Colaresi. 2020. "Human Rights Are (Increasingly) Plural: Learning the Changing Taxonomy of Human Rights from Large-Scale Text Reveals Information Effects." *American Political Science Review* 114 (3): 888–910. DOI:10.1017/S0003055420000258.
- Park, Baekkwon, Amanda Murdie, and David R. Davis. 2019. "The (Co)Evolution of Human Rights Advocacy: Understanding Human Rights Issue Emergence over Time." *Cooperation and Conflict* 54 (3): 313–34. <https://doi.org/10.1177/0010836718808315>.
- Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson, and David G. Rand. 2014. "Structural Topic Models for Open-Ended Survey Responses." *American Journal of Political Science* 58 (4): 1064–82. <https://doi.org/10.1111/ajps.12103>.
- Schmidt, Brian C. 2016. *The Political Discourse of Anarchy: A Disciplinary History of International Relations*. New York: State University of New York Press.
- Schönfeld, Mirco, Steffen Eckhard, Ronny Patz, and Hilde van Meegdenburg. 2019. "The UN Security Council Debates 1995–2017." *arXiv preprint arXiv: 1906.10969*. <https://arxiv.org/abs/1906.10969>.
- Schuler, Paul. 2020. "Position Taking or Position Ducking? A Theory of Public Debate in Single-Party Legislatures." *Comparative Political Studies* 53 (9): 1493–524. <https://doi.org/10.1177/0010414018758765>.
- Slapin, Jonathan B., and Sven-Oliver Proksch. 2008. "A Scaling Model for Estimating Time-Series Party Positions from Texts." *American Journal of Political Science* 52 (3): 705–22. [www.jstor.org/stable/25193842](http://www.jstor.org/stable/25193842).
- Smith, Courtney B. 2006. *Politics and Process at the United Nations: The Global Dance*. Boulder, CO: Lynne Rienner.
- Steffek, Jens, Marcus Müller, and Hartmut Behr. 2021. "Terminological Entrepreneurs and Discursive Shifts in International Relations: How a Discipline Invented the 'International Regime.'" *International Studies Review* 23 (1): 30–58. <https://doi.org/10.1093/isr/viaa003>.
- Titunik, Roció. 2015. "Can Big Data Solve the Fundamental Problem of Causal Inference?" *PS: Political Science & Politics* 48 (1): 75–79. <https://doi.org/10.1017/S1049096514001772>.
- Watanabe, Kohei, and Yuan Zhou. 2020. "Theory-Driven Analysis of Large Corpora: Semi-Supervised Topic Classification of the UN Speeches." *Social Science Computer Review* 40 (2): 346–66. <https://doi.org/10.1177/0894439320907027>.
- Whyte, Christopher. 2019. "Can We Change the Topic, Please? Assessing the Theoretical Construction of International Relations Scholarship." *International Studies Quarterly* 63 (2): 432–47. <https://doi.org/10.1093/isq/sqy050>.
- Wiener, Antje. 2007. "The Dual Quality of Norms and Governance beyond the State: Sociological and Normative Approaches to 'Interaction.'" *Critical Review of International Social and Political Philosophy* 10 (1): 47–69. <https://doi.org/10.1080/13698230601122412>.
- Wilkerson, John, and Andreu Casas. 2017. "Large-Scale Computerized Text Analysis in Political Science: Opportunities and Challenges." *Annual Review of Political Science* 20:529–44. <https://doi.org/10.1146/annurev-polisci-052615-025542>.