Received: 6 April 2023 | Revised: 14 February 2024 | Accepted: 13 March 2024

DOI: 10.1111/psj.12534

ORIGINAL ARTICLE

PSJ POLICY STUDIES JOURNAL

Staying on the democratic script? A deep learning analysis of the speechmaking of U.S. presidents

Amnon Cavari[1] | **Akos Mate**[2] | **Miklós Sebők**[2]

[1]Lauder School of Government, Diplomacy and Strategy, Reichman University, Herzliya, Israel

[2]Department for Government and Public Policy, Centre for Social Sciences, Budapest, Hungary

Correspondence
Amnon Cavari, Reichman University, 8 HaUniversita St., P.O.Box 167, Herzliya, 46150, Israel.
Email: cavari@runi.ac.il

Funding information
Artificial Intelligence National Laboratory of Hungary, Grant/Award Number: 101008468; European Union, Grant/Award Number: RRF-2.3.1-21-2022-00004; Hungarian Academy of Sciences: V-Shift Momentum Project; Israel Science Foundation, Grant/Award Number: 1271/18

Abstract

Dynamic agenda representation assumes a linkage between the policy emphases prescribed by various democratic inputs (electoral promises and public opinion polls) and policy agendas ranging from the media to executive orders. An extrapolation of this idea would propose that, in the U.S. context, policy emphasis in major programmatic messages such as State of the Union addresses would be followed by the president's day-to-day communication. We investigate this congruence with a new database of presidential speeches that, for the first time, offers a deep learning-enhanced sentence-level policy topic coding of various forms of the speeches U.S. presidents made from Truman to Trump (for a total count of 16,523 speeches divided into nearly 2 million individual sentences). Using this database, we demonstrate that presidents' occasional, day-to-day remarks strongly correlate with the annual policy messages—in this sense, presidents are staying on the democratic script.

KEYWORDS

content analysis, deep learning, large language models, policy agendas, state of the union addresses

The mandate theory of representation assumes a normatively desirable connection between the transparency of elected officials' pledges and how they follow up on their pledges and implement them while in office. In their theory of representation, Manin et al. (1999) combine this approach with accountability theory, which is based on the notion that politicians want to deviate from the electoral mandate in the "public interest" and convince the electorate in the next election that, indeed, they acted on their promises. At this point, representation becomes dynamic and intertemporal, with reputation and persuasion becoming critical elements of the process of representation, which moves beyond mandate theory and its focus on singular elections as moments where representation is established.

Similarly, Bevan and Jennings (2014) put forth a theory of dynamic agenda representation, which combines elements of 'promissory' and 'anticipatory' representation, where promissory is tied to mandates, and anticipatory conveys the idea of dynamic responsiveness and accountability. They situate

executive speeches in this general framework as an area where dynamic representation takes place in major executive addresses that spell out policy content and promises while in office. These executive addresses are used to communicate general priorities and, therefore, perform a substantive function in the representative process.

An extrapolation of the idea of speech-to-policy in executive politics would be that policy emphases of heads of government in major speeches should be followed by day-to-day communication. If they do, heads of government can be viewed as staying on the democratic script.

In the United States, the pivotal programmatic speech is the president's State of the Union (SOU) address delivered each year to Congress. It is widely broadcast and watched by the American people. The annual speech serves presidents as an opportunity to restate the commitments they laid down in their inaugural messages, to revive the principles to which they committed their presidencies, and to show how those principles will be reflected in their legislative programs (Campbell & Jamieson, 2008; Hoffman & Howard, 2006).

Beyond programmatic messages, presidents constantly engage with the American people by delivering frequent messages to various audiences at public events (Cavari, 2017). What is the relationship between the high-profile programmatic speeches of the State of the Union (SOU) and the routine or day-to-day public speeches of presidents?[1] Are presidents using the SOU addresses to set a policy agenda that they then follow with their routine public remarks, or do presidents see these public remarks as a separate governing (or electoral) tool? In other words, are presidents staying on the democratic script in their day-to-day public remarks?

In this article, we answer these questions by offering a three-fold contribution. First, we propose a novel theoretical explanation, related hypotheses, and operationalization to study programmatic and day-to-day presidential communication congruence. Second, for the first time, we apply a deep learning-supported sentence-level policy topic coding for a comprehensive dataset of American presidential speeches. Third, our analysis demonstrates a strong correlation between the presidents' SOU programmatic annual messages and their occasional remarks (OR)—in this sense, we argue that presidents are staying on the democratic script.

DYNAMIC AGENDA REPRESENTATION AND THE DEMOCRATIC SCRIPT

The mandate theory of representation assumes a program-to-policy linkage of representative democracy: policymaking should follow the emphases of electoral manifestos and other policy pledges. This linkage is a normatively desirable connection because it creates a transparent "contract" between the electorate and elected officials.

There are two major methodological strategies for measuring the fulfillment of such mandates: the pledge and the saliency approach (Körösényi & Sebők, 2018; Royed, 1996). Pledge-based approaches focus on identifying individual promises (rhetoric or policy-oriented) and how the incoming administration follows up on them. An often-cited case of a pledge gone wrong is George H. W. Bush's "read my lips" speech on how there would be no new taxes. Saliency-based approaches concentrate on the overall distribution of policy emphases. For example, if an electoral manifesto is packed with references and mentions of social policy, the expectation is that budget appropriations should follow a similar trend (Budge & Hofferbert, 1990). Both approaches focus on the program-to-policy linkage (Thomson, 2001) and suggest a clear trail of how a democratic mandate works in practice. As long as politicians deploy adequate policy emphases (or pay attention to concrete or less concrete pledges as signals for these emphases), they are staying on what can be called the democratic script.

In their seminal work on democratic representation, Manin et al. (1999, p. 29) state that "under democracy, governments are representative because they are elected," and the "winning platform becomes the mandate that the government pursues." Yet, politicians are not limited to their campaign promises but, instead, want to deviate from the electoral mandate in the "public interest" and

convince the electorate in the next election that they did so. At this point, representation becomes dynamic and intertemporal, with reputation and persuasion becoming critical elements of the process of representation. Such a model of representation moves beyond the focus on elections as moments where representation is established to a model that incorporates various entry points to adjust and change policy promises.

Bevan and Jennings (2014) make a similar argument when they put forth their theory of dynamic agenda representation, which combines elements of what Mansbridge (2003) calls 'promissory' and 'anticipatory' representation. The former fulfills the mandate promise, whereas the latter conveys the idea of dynamic responsiveness and accountability. Bevan and Jennings situate executive speeches in this general framework as one of the three areas where dynamic representation takes place. The head of government delivers an annual formal statement setting out his or her policy priorities for the year ahead. These speeches are forward-looking, communicating general priorities and specific measures the executive intends to address next year. This substantive function of executive speeches are then, or should be, reflected in their transmission into policy outcomes.

An extrapolation of this speech to policy linkage to executive speechmaking would propose that in day-to-day communication, heads of government would adhere to the policy emphases they made in major speeches that spelled out their policy content and promises. Building on the literature on dynamic agenda representation, we flesh out a novel theoretical proposition related to the political communication of U.S. presidents—assessing the congruence between major programmatic messages of presidents and their daily communications with the American people. For presidents, speeches are leadership tools used to demonstrate their commiment to an issue to the American public, to members of Congress, and to the executive branch (Cavari, 2017; Cohen, 1995, 2008, 2012; Eshbaugh-Soha, 2016; Eshbaugh-Soha & Juenke, 2022; Heith, 2013). As long as presidents stay on message, they are officeholders who demonstrate that they stay faithful to their promises to the electorate and stay on the democratic script. If they do, they further demonstrate to the public that they are focused on policy and are making an orchestrated effort to set the agenda and to persuade the public of the advantages in supporting their policy proposals (Cavari, 2017; Kernell, 2007; Rottinghaus, 2010).

THE ROLE OF PRESIDENTIAL SPEECH TYPES IN DYNAMIC AGENDA REPRESENTATION

Repetition and "retail" sales of policy ideas and decisions are part and parcel of how presidents convince the electorate that they comply with their pledges and the overall emphases of their electoral mandate (Cavari, 2017; Cohen, 2010). As circumstances change, presidents update and elucidate their policy agenda in major speeches, which serve as a bridge between them and the electorate. The tool most strongly used by presidents to reiterate their commitments to the American people is the State of the Union (SOU) address. These are highly visible policy manifestos used by presidents to present their programmatic priorities to Congress and the American people (Campbell & Jamieson, 2008; Light, 1999; Yates & Whitford, 2005).

The SOU serves presidents an opportunity to restate the commitments they laid down in their inaugural messages, to revive the principles to which they committed their presidencies, and to show how those principles will be reflected in their legislative and executive programs (Campbell & Jamieson, 2008). Presidents use these highly visible events (Barabas, 2008; Baum & Kernell, 1999) to advocate for policies already being considered by Congress, to introduce innovative ideas, or to threaten vetoes (Hoffman & Howard, 2006), and they usually follow up on these speeches with policy action (Cummins, 2008; Lovett et al., 2015). As such, SOU constitutes the ideal type for the annual reiteration of the electoral mandate adjusted to a new reality in office and to itemize the steps to be taken to realize these promises. President Clinton articulated well the idea of the SOU as a programmatic promise in his assessment of his time in office: "We used every State of the Union address to plan a specific detailed agenda for the coming year, and then we just rammed it and worked on it."[2]

Beyond the well orchestrated, high attention State of the Union addresses, presidents constantly engage with the American people, delivering speeches almost daily (Cavari, 2017; Cohen, 2008, 2010; Eshbaugh-Soha, 2016). How do the major programmatic speeches of the State of the Union relate to the day-to-day communication efforts of presidents? We define a speech as a public event in which the president gives an oral remark. The definition of a speech covers all such events, from major national addresses to occasional remarks on the tarmac before boarding Air Force One. This broad definition covers all oral remarks, regardless of audience size: from the general audience of a live network broadcast to speeches in small town-hall-style events with little to no live TV or radio coverage. Altogether, these speeches portray the presidents' continuous, day-to-day routine and diverse means of connecting with the American people.

In Figure 1, we plot the number of speeches per week of each president from Truman to Trump ($N = 22{,}154$). The data include all oral remarks of presidents from Truman through Trump, which we downloaded from The American Presidency Project, a widely used online archive of the public papers of presidents (https://www.presidency.ucsb.edu).

During the second half of the 20th century, U.S. presidents have increasingly appealed to the American public. While the overall upward trend from Truman to Clinton has recently reversed, presidents still deliver, on average, nearly one speech per day. This frequency of public events speaks to the dominant role of presidents in shaping the agenda of U.S. politics (Cavari, 2017). The increase in the number of public remarks reflects a rising need to develop a direct connection with the public, a need that is fueled by the technological advances of radio, television, and, more recently, the internet (Cohen, 2008, 2010; Fontaine & Gomez, 2020; Haeder & Chattopadhyay, 2022; Heith, 2013, 2020; Ouyang & Waterman, 2020). To aid them in this task, presidents have developed a sizeable institutional apparatus constantly working to deliver the president's message (Kumar, 2007). This institutional apparatus allowed presidents to make frequent remarks not limited to major orchestrated or mediated events that are part of presidents' daily governing activities.

Presidents' speeches range from major required events such as speeches to a joint session of Congress, Inaugural speeches, or the fully orchestrated major to the nation speeches, to the more frequent routine, nearly day-to-day occasional events. The latter category includes speeches that are usually at the president's discretion, have little expected guidelines, and offer the president the opportunity to address any issue he desires as he attunes to the political circumstances ($N = 16{,}450$).

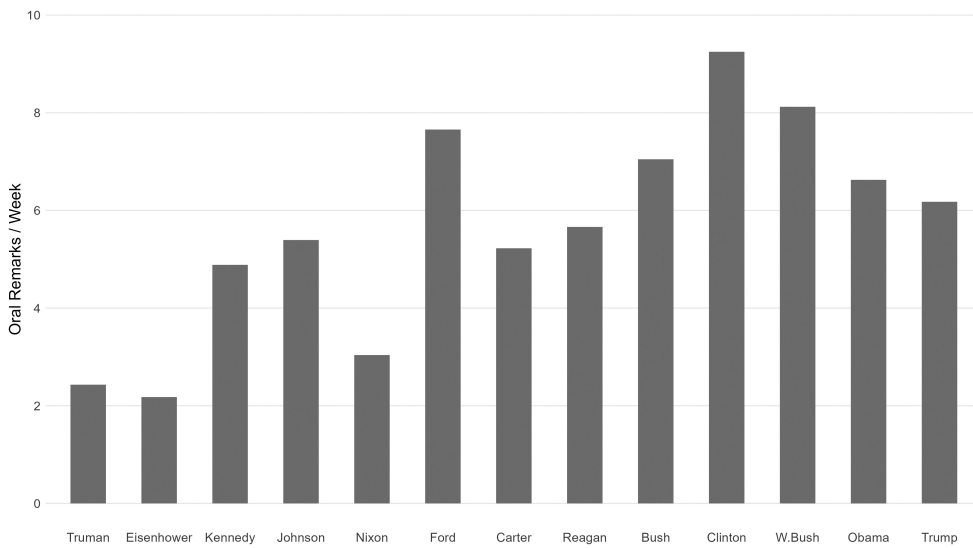

FIGURE 1 Number of speeches a week from Truman to Trump.

For all presidents since Truman, occasional remarks (OR) have occupied the largest share of their speech-making efforts (mean = 75%, *SD* = 6.58). While we see a gradual increase in the number of speeches over time (Figure 1), we find no clear time series trend in the share of occasional remarks, suggesting that occasional remarks occupy a relatively stable share of the presidents' public appeals.[3] These speeches, therefore, offer us an opportunity to assess the connection between the institutionalized annual policy pledge—the SOU—and the routine day-to-day public connection with the public.

There is a good reason to expect that presidents follow up on their programmatic annual message with daily, occasional public remarks. Presidents understand the importance of constantly campaigning for their message as part of their governing strategy (Blumenthal, 1980; Gergen, 2008; Ornstein & Mann, 2000) and devote an increasing amount of time and resources to their public relations (Kumar, 2007) to manage their agenda and to appeal to various audiences (Cavari, 2017; Cohen, 2010; Heith, 2013; O'Brien, 2019). These public remarks serve as a leadership tool for presidents to demonstrate that they are responding to public demands and are engaged in and lead the policy agenda (Eshbaugh-Soha, 2016; Eshbaugh-Soha & Juenke, 2022; Villalobos et al., 2012; Whitford & Yates, 2009). In this view, presidential routine remarks are not used to pay lip service to public interests but are orchestrated efforts to engage with the public to mobilize public support for their policy preferences, which, in turn, can affect policy making (Canes-Wrone, 2006; Eshbaugh-Soha & Miles, 2011).

In their routine public appeals, presidents mix audience sizes and formats. They speak in small events to dedicated audiences, to large crowds, or even just deliver a few words to a local or national television channel. Consider President Obama's speeches following the State of the Union on January 12, 2016. While the president spoke at length at the Capitol, he followed up on his address with a series of public events where he emphasized parts of his message—giving speeches, among others, to high school students in Baton Rouge, Louisiana, university students at the University of Nebraska, political leaders in Baltimore, Maryland, workers of the auto industry in Detroit, and federal workers in the nation' Capital. The president was very clear about the importance of this tour, saying to his audience in Nebraska the day after the SOU address: "But whenever I give a State of the Union, I want to get out of Washington and talk to people out in the country."[4]

In a series of studies spanning presidents from Reagan to Trump, presidential scholar George Edwards questions the effect of these speeches on public opinion and its success as a governing strategy (Edwards, 2003, 2009, 2012, 2016, 2021). Various studies challenge this notion and suggest a more substantial effect of presidential speeches. These studies demonstrate that the political environment conditions the effect of the presidents' messages (Cavari, 2013; Rottinghaus, 2010) and that presidents are more successful in pushing their agenda when they make repeated appeals (Barrett, 2004; Cavari, 2017; Fett, 1992, 1994; Wood, 2007) to various audiences (Cohen, 2010; Heith, 2013).

Of course, beyond repeating strategic policy priorities, occasional remarks can also serve as testing grounds for new ideas. Since the stakes are smaller, ORs may serve as a tool to refine policies before they reach the highly publicized environment of the SOU. While such alternative uses may be common, we argue that these routine speeches are a leadership tool that presidents use to advance their agenda, advertise their actions, and persuade voters of their credibility and trustworthiness (regardless of the original mandate). As such, we propose that routine speechmaking can be used to assess the connection between programmatic messages (SOU) and day-to-day policy focus and political action. In brief, daily communication is also part of what the presidents "do" to maintain the representational nature of democracy.

In light of these considerations, we put forth four hypotheses related to the role of presidential speeches in an extended theory of dynamic agenda representation (the last two of which are connected). First, we expect that presidents would follow up on their ceremonial high-profile policy speech each year with repeated appeals to the American people during the remainder of the year. That is, we expect the policy emphases in the SOU address each year will be positively correlated with the similar emphases of the presidents' occasional remarks that year. We can, therefore, formulate our first hypothesis as follows:

H1. The policy agenda of the most important programmatic speech (SOU) and of routine remarks (ORs) each year will be positively correlated.

And yet, we expect presidents to attend to evolving issues that affect the public agenda. We therefore expect a declining attention to issues that were dominant at the time of the State of the Union address.

H2. The correlation between the policy agenda of the most important programmatic speech (SOU) and that of the routine remarks (ORs) will steadily decline over the course of the year.

Finally, the president's agenda is affected by events, some of which may overwrite or alter their intended agenda as presented in the SOU. For example, in his first State of the Union, in January 1990, President George H.W. Bush discussed the changing world order following the collapse of the Soviet Union yet focused most of his attention on domestic issues and investments. In August, however, Saddam Hussein invaded Kuwait, which triggered an American response that defined the president's focus until the end of his term.

We do not expect that an event, or a series of events, will alter the policy attention to all issues. Instead, we expect that events—most of which address core functions of government—will affect the distribution of attention among the different policies—shifting more attention to one or a few issues at the expense of other issues. Hence, our third hypothesis focuses on the effect of events on the diversity of attention—increasing attention to one or a few issues at the expense of others, which would decrease the diversity of policy attention (Jennings, Bevan, Timmermans, et al., 2011). We further suggest that this would be conditioned on the diversity of attention in the SOU. When presidents have a broad programmatic agenda (high diversity), we should expect that events would have a more substantial effect on narrowing the agenda. When presidents already have a narrow agenda, the effect would be negligible.

H3a. Major domestic and foreign events decrease the diversity of the presidents' routine attention (measured in ORs) relative to that presented in strategic communication (based on SOU).

H3b. The effect of domestic and foreign events on the diversity of routine agenda would be conditioned on the diversity of the annual agenda in the SOU.

As for the coherence of these three hypotheses, the respective pairs of H1 with H2 and H1&H2 with H3 pose no issues. The first pair, collectively, states that from a high congruence starting point between the two agendas, there is a diminishing rate of congruence over the rest of the year. H1 and H2 refer to policy attention, whereas H3 relates to the diversity of attention.

THE MEASUREMENT OF THE POLICY AGENDA OF PRESIDENTIAL SPEECHES

Scholars take various approaches to analyze presidential addresses and infer from them about presidents' policy agenda or attention. Research designs vary in the scope of data analyzed, the required unit of observation for analysis, and the type of inference made from the data. Most studies examine a select number of speeches focusing on specific attention to a few policy domains (Cavari, 2017; Cohen, 1995; Cummins, 2008). Other studies seek to assess the broad agenda of the president and, therefore, adopt an expansive approach examining a more inclusive set of speeches but making a relatively general reference to specific issue interests such as foreign policy or the economy (Eshbaugh-Soha & Peake, 2004, 2005; Peake, 2001; Wood, 2007; Wood & Peake, 1998). Neither of these approaches is suitable to our research

question, which concerns the congruence of the various policy topics in programmatic speeches and occasional remarks.

The Comparative Agendas Project (CAP) offers a suitable and widely used coding scheme for policy topic distribution (Bevan, 2019). The scheme uses a detailed codebook of major policy topics divided into subtopics to gauge the distribution of policy emphases in different datasets, from public opinion poll questions to legislative and executive actions. Executive speeches—including U.S. presidential communication—are an inherent part of this collective effort. For example, Jennings and colleagues use humanly coded primary speeches of heads of the executive in six countries to assess the variation in issue attention in these speeches (Jennings, Bevan, Timmermans, et al., 2011). Several studies use this toolkit to evaluate U.S. presidential speeches more closely (Cavari, 2017; Cavari et al., 2021; Russell & Eissler, 2022).

In assessing the policy attention in speeches of cheif executives, the majority of existing work taking a CAP approach applies a sentence-level research design (Bevan & Rasmussen, 2020; Jennings, Bevan, & John, 2011; Jennings, Bevan, Timmermans, et al., 2011; Jennings & John, 2009; Mortensen et al., 2011; Russell & Eissler, 2022). These studies focus on the distribution of attention to various policies—saliency of policies and diversity of attention—rather than examine the overall attention to one or more policy domains. Therefore, sentences are the most reasonable unit of analysis for proper analysis of policy attention in speeches, which may be wide-ranging in terms of topic coverage even in briefer formats. Such a research design also allows a natural weighting scheme that transcends particular events and rhetoric styles (Wood, 2007).

Setting the unit of observation on the sentence level opens a new methodological challenge. Coding a corpus of thousands of presidential speeches from 1945 through 2020 at the sentence-level, is an enormous task. Millions of sentences from thousands of speeches with different event characteristics (function, length, number, and diversity of policy topics covered) must be processed with high accuracy, reliability, and adequate cost. At this scale, human labor is increasingly substituted for machine learning tools; tools that are quickly becoming a new standard in social research (Loftis & Mortensen, 2020; Sebők et al., 2022; Sebők & Kacsuk, 2021).

A rich body of work relies on traditional machine learning algorithms, such as random forest or support vector machine. More recent work relies on the rapidly growing pre-trained large language models that show superior results in terms of coding accuracy for a wide variety of tasks (Devlin et al., 2018; Liu et al., 2019). The breakthrough moment was the development of the transformer architecture that allows unsupervised learning for training (thus removing the need for task-specific training) (Vaswani, 2017). These models, trained on a massive amount of public text data, can then be fine-tuned with domain-specific data for a fraction of the labor and time cost of human coding (Wankmüller, 2022). One of the earliest models to rely on transformers was the BERT (Bidirectional Encoder Representations from Transformers), which paved the way for further variants (Devlin et al., 2018).

As language models can rely on an ever-increasing size both in terms of training data and parameters, researchers have multiple options to choose models that match the underlying research task. A particularly suitable option for our research design is RoBERTa, which stands for "robustly optimized BERT pretraining approach" (Liu et al., 2019). RoBERTa contains 355 million parameters and is pre-trained on 160GB of English language text. This model improves the performance of the initial BERT model by using a different set of pretraining choices and a considerably larger training set. This more extensive, English-language-focused model serves as a suitable platform for the extensive multi-class classification necessary for CAP-style tasks and for the relatively in-formal language that characterizes many occasional remarks of presidents.

DATA AND METHODS

We examine the congruence of policy agendas between programmatic speeches and occasional speeches with a multi-layered research design. The dominant elements of the pipeline include data collection/

cleaning, deep learning supported coding on the sentence level for the 20 CAP topics used in the U.S. project (Jones et al., 2023),[5] and a panel regression for hypothesis testing purposes.

Starting with data collection, we downloaded our initial dataset from the American Presidency Project (APP) online database covering all SOUs and occasional remarks of President Truman through President Trump. For each dataset, we used the categorization of APP. The SOU dataset covers 73 speeches with 20,595 sentences.[6] Occasional remarks are based on the category of documents in the APP with the same title.[7] This includes 16,450 speeches, divided into 1,794,662 individual observations for coding.

We used a large language model trained on human-coded data to code our data into the 20 CAP policy categories. Due to the relatively uncharted waters of using large language models for CAP coding, we applied a five-step model development and data labeling workflow (see Figure 2). This model was inspired by the algorithm-agnostic proposal in Sebők and Kacsuk (2021), which allows for an expert-in-the-loop (also called active learning) approach that can help to iteratively balance the training data (a critical step for handling unbalanced datasets).

In the first step of our workflow, we fine-tuned the large language model using the double-blind coded training data set based on a stratified sample of presidential speeches containing 51,390 sentences from 466 occasional remarks. With the next stage in mind—machine coding of sentences—coders were instructed to assess each sentence independently without accounting for the context of the speech. In our coding, we also follow the CAP approach that refers to mere attention, quantified by the count of mentions. In doing so, we do not account for the message's concreteness, the message's frame, the direction of the message, or an interpretation of how it is mentioned in the context of the speech (Baumgartner et al., 2019).

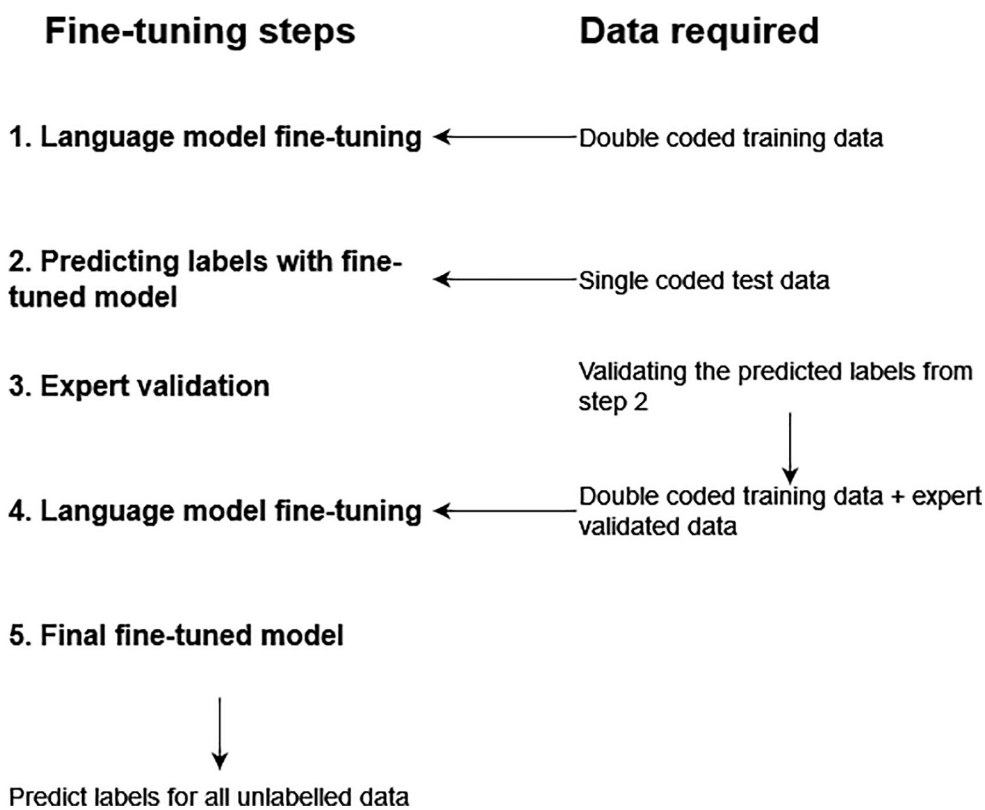

FIGURE 2 The human-in-the-loop model development and classification workflow.

Our focus on mentions without accounting for context improves the quality of the data and avoids subjective interpretation. We also do not differentiate between speeches in terms of exposure and audience size. Such data are not available, and, if they were, it is not clear that they will allow us to assess the impact of attention in a speech efficiently. A presidential mention of a topic in a small, brief unattended remark can have a stronger public impact than a mention of a topic in a large, widely broadcasted speech. Therefore, we refer to all policy mentions, regardless of the audience and location of the speeches.

The level of agreement among human coders was at least 87% and Conger's Kappa at least 0.73 ($p < 0.00$) for each president. We experimented with various parameter specifications through extensive hyperparameter testing for both BERT and RoBERTa models, following the hyperparameter recommendations in Devlin et al. (2018). The results were robust across the hyperparameter space, and RoBERTa offered better performance over BERT; therefore, we decided to retain the former for the rest of the workflow.[8]

In the second step, we used the fine-tuned RoBERTa model to predict labels for the single-coded sample of 363,517 sentences (from a randomly assigned 3364 occasional remarks clustered by presidents). This allowed us to compare the accuracy of our model to the human-coded data. In the third step, we validated the prediction results using expert coders who served as tiebreakers in the cases of disagreements between machine and expert coding (we focused only on predictions for policy categories). This step in the workflow is a crude approximation of active learning. In active learning designs, the machine learning algorithm queries the expert for labels. Instead of an automated solution, in the proposed workflow, the researcher can make an informed decision on the focus of the validation efforts.

This active learning element also allows us to target specific categories with low counts in the training data. The unbalanced data issues can be ameliorated by focusing expert validation efforts on the relatively rarer categories. In this third step, we obtained 8142 expert-validated labels added to our training data. As a fourth step, we again fine-tuned the RoBERTa model using this enhanced training data containing the original double-coded training data and the expert-validated data from the third step. This means the final fine-tuned model is trained on the double-coded sentences and the additional 8142 expert-validated sentences (59,532). With the model finalized in Step 5, we labeled both SOUs and the remaining uncoded set of occasional remarks (1,379,755 sentences) with the model in Step 6.

The data generated provides the first large-scale, systematic assessment of the policy agenda of presidential speeches we are aware of. It covers 1,794,662 sentences of occasional remarks, of which 76.88% (1,379,755) are machine-coded, and 20,595 sentences of SOU (100% machine coded). Table 1 summarizes the coding source in our final dataset—combining a human-coded training set, human-coded validation data, and data solely machine-coded by our RoBERTa model.

Of the coded sentences in the SOU and occasional remarks, 7067 and 1,178,717 did not contain policy information (none of the 20 CAP major topic codes assigned). In what follows, we only focus on the sentences with policy contents (13,528 (66%) and 615,945 (34%), respectively). The difference in policy-labeled sentences conforms with what we can expect of the difference between the policy-focused SOU addresses (Campbell & Jamieson, 2008; Hoffman & Howard, 2006) and the more flexible and routine occasional remarks that serve presidents with multiple leadership tasks.

TABLE 1 Composition of the final sentence-level CAP-coded dataset by source.

Policy label source	OR sentences coded	SOU sentences coded
Human coder: double-blind	51,390	
Human coder: single	363,517	
Machine: RoBERTa	1,379,755	20,595
Final dataset	1,794,662	20,595

Figure 3 summarizes the distribution of the policy sentences of the final dataset of OR speeches into the 20 CAP categories. The figure reveals the unbalanced nature of the data. Two policy categories—international affairs and macroeconomics—occupy about 40% of the policy agenda. Five policy categories—defense, government operations, health, education, and law and crime—cover 5% to 8%. Another five categories—civil rights, welfare, foreign trade, energy, and domestic commerce—have a share of around 3%, and the eight remaining topics received less than two percent of the overall presidential policy attention.

We also performed several validation and robustness checks for various pipeline steps. First, in supplementary online Appendix A, we compare the final classification results of the human-in-the-loop, active learning-enhanced workflow summarized in Figure 3 with the model that does not use the expert-validated data for fine-tuning (steps 1–5 vs. just steps 1 and 5). While overall accuracies are similar, the active learning workflow produced better results for low sample size categories—a crucial limitation of using out-of-the-box models for labeling unbalanced datasets. Second, in supplementary online Appendix B, we benchmark the fine-tuned RoBERTa model against commonly used supervised machine learning approaches. Our results show a clear superiority of the language model approach vis-á-vis traditional machine learning, with the headline weighted F1 score showing a big gap of 0.18 versus the best-performing machine learning algorithm (Support Vector Machine).

Third, in supplementary online Appendix C, we compare the machine-classified SOU dataset to a similar, human-coded data source from the Policy Agendas Project. The fine-tuned language model closely tracked the expert-coded results for 16 of the 20 major topics. Only four major topics provided significantly different shares with non-overlapping margins of errors, and differences shrinks once we combine the neuralgic topics of defense and foreign affairs.[9] Overall, the deep learning approach yields comparable results with little marginal costs after fine-tuning the model for American presidential speeches. Finally, we examined the added accuracy of our model when we included the human-coded dataset of the SOU addresses. We see no change in all benchmarks of our RoBERTa model following this test. This further demonstrates the quality of our training set.

With the hybrid (human and machine-coded) validated dataset at hand, we proceeded to the quantitative analysis of our hypotheses. We first illustrate the congruence of agendas by plotting the correlation between the two series—SOU and OR—for each topic. We add to that with an analysis of agenda

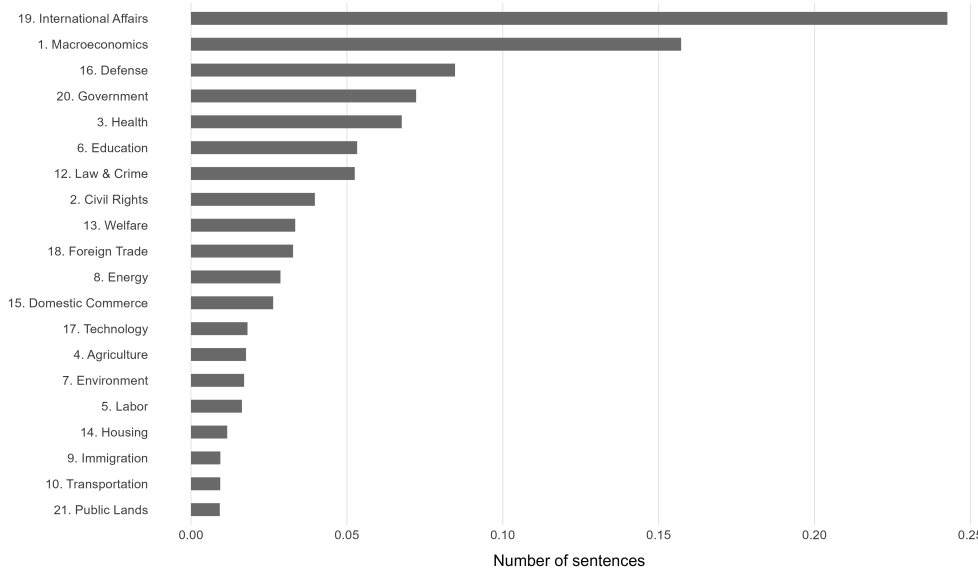

FIGURE 3 Policy (CAP) distribution of the final dataset of occasional remarks.

stability (Mortensen et al., 2011)—measuring the average stability of the agenda between SOU each year (y) and OR in each quarter of that year (yq).[10]

To test Hypothesis 1, we used a panel time series regression strategy to calculate the relationship between policy emphasis in the annual State of the Union address and the average emphasis over the rest of the year in the occasional remarks by major topics. For Hypothesis 2, we applied the same model, divided into the four calendar quarters to examine the declining effect of SOU attention over time.

For Hypothesis 3, we calculated the diversity of quarterly rhetorical agenda using the Shannon H entropy score commonly used in CAP research (Boydstun, 2013; Cavari et al., 2021, 2023; Jennings, Bevan, Timmermans, et al., 2011). To account for events, we used the "ordinary domestic and international events" dataset by Newman and Forcehimes (2010), as updated by Ostrom et al. (2021). The measure is a count of events in each quarter that were covered at least three days a month on the front page of the *New York Times*. Because the events-data are from 1953, we use data from 1953 to 2000 to test hypothesis 3.

This research design yielded the quarter-topic database structure presented in Table 2. For each quarter, we have 20 observations, one for each CAP category—totaling 5840 quarter-topic observations (5768 with events data). Our variables include the president, percent SOU (annual), percent OR (aggregated to quarters), entropy score of SOU (annual) and OR (quarter), and number of events.

RESULTS

Figure 4 summarizes the overall correlation between the policy emphases of SOU and OR. The correlation is high for most topics. Seventeen topics are strongly correlated, ten with a coefficient higher than 0.5 (high), and seven with a correlation above 0.3 (moderate). The overall weighted average correlation is 0.5. Applying dickey fuller tests to all series—SOU and OR for each topic—reveals that all series are stationary. The average agenda stability also yields substantial congruence between SOU and OR. Starting at an average stability of 0.72 (or 72% match) in the first quarter each year and declining to 0.69 in the last quarter. Both initial tests lend initial support for H1.

To test the independent effect of presidential emphasis in SOU on OR, we estimated a regression model where the dependent variable is defined as percent attention in occasional remarks to a topic in a quarter, and the main explanatory variable is presidential emphasis in the SOU. To account for the variation between presidents and the variation caused by the political calendar, we included fixed effects for presidents (13 presidents) and quarters (1 through 4) and clustered the errors by presidents. Finally,

TABLE 2 The structure of the database.

Year	Quarter	Major topic	President	SOU CAP%	OR CAP %	OR diversity	SOU diversity	Events
1953	Q1	Macroeconomics	Eisenhower	20	26	1.91	2.35	1
1953	Q2	Macroeconomics	Eisenhower	20	8	1.88	2.35	1
1953	Q3	Macroeconomics	Eisenhower	20	13	1.90	2.35	2
1953	Q4	Macroeconomics	Eisenhower	20	4	2.13	2.35	3
1953	Q1	International affairs	Eisenhower	16	24	1.91	2.35	1
1953	Q2	International affairs	Eisenhower	16	42	1.88	2.35	1
1953	Q3	International affairs	Eisenhower	16	33	1.90	2.35	2
1953	Q4	International affairs	Eisenhower	16	37	2.13	2.35	3
1953	Q1	Agriculture	Eisenhower	4	4	1.91	2.35	1
1953	Q2	Agriculture	Eisenhower	4	1	1.88	2.35	1
1953	Q3	Agriculture	Eisenhower	4	5	1.90	2.35	2
1953	Q4	Agriculture	Eisenhower	4	16	2.13	2.35	3

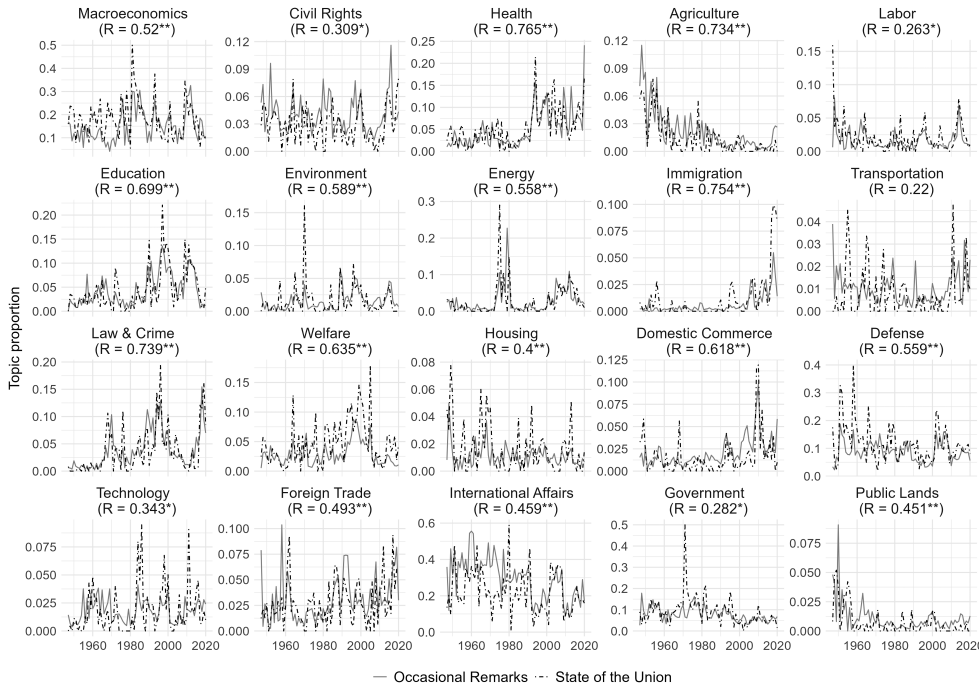

FIGURE 4 Correlations between speech types by policy topic.

TABLE 3 The effect of SOU policy emphases on OR policy emphases.

	1	2	3	4	5
DV: OR policy emphasis (Q/topic)	Overall	Q1	Q2	Q3	Q4
SOU policy emphasis$_t$	0.367***	0.463***	0.405***	0.330**	0.270**
	(5.06)	(6.23)	(5.93)	(3.60)	(3.60)
Fixed-effects					
President	Yes	Yes	Yes	Yes	Yes
Quarters	Yes	No	No	No	No
Major topic	Yes	Yes	Yes	Yes	Yes
Observations	5840	1460	1460	1460	1460
R^2	0.699	0.720	0.745	0.651	0.698

Note: Standardized beta coefficients; *t* statistics in parentheses. Errors are clustered by presidents.
*$p<0.05$; **$p<0.01$; ***$p<0.001$.

considering the variation between topics, as illustrated in Figure 4, we added fixed effects for topics (CAP categories). In sum, our model could be summarized using Equation (1):

$$\mathrm{OR}_{t,q} = b_1\mathrm{SOU}_{t,y} + b_2\mathrm{President}_j + b_3\mathrm{Quarter}_q + b_4\mathrm{MajorTopic}_m + \in_j \quad (1)$$

where t is the major topic (1 to 21, 11 is missing), q is the quarter (January–March, April–June, July–September, October–December), y is the year (1947–2020), j is president (Truman, 33 through Trump, 45), and m is CAP major topic category (1–21). The results are summarized in column 1 of Table 3.

The results lend support for Hypothesis 1. The policy emphasis of a president in the State of the Union is positively associated with policy emphasis in occasional remarks. A change of one standard deviation in SOU policy emphasis increases presidential attention to the issue by 0.367 standard deviations.

For robustness check, we estimated the model with lagged terms for OR (1 quarter). While SOU's effect size is now smaller (Beta = 0.218, $t = 4.80$), as can be expected when including lagged measures of quarterly attention, the effect of SOU at the beginning of the year on daily attention is still positive and significant. We also tested the robustness of these models by including per category and total sentence counts for occasional remarks and State of the Unions without any significant change in the results and estimated effects. Similarly, we have not found any significant effect for the count of ORs per quarter. We estimated a model using topics as random effects. The LR test indicates a better fit than the linear model, but point estimates are relatively similar. In the interest of simple interpretation, we present here the results of the basic linear model. Finally, we address the concern that the results are driven by a few large topics that capture most of the presidents' attention—mainly economy and foreign affairs (including defense and international relations). Estimating the model without these topics results in the same (and slightly higher) effects (Beta = 0.418, $t = 5.18$).

In Figure 5, we plot the marginal effect by topic. For each topic, we used the margins for minimum attention in SOU, mean attention and maximum attention of the topic. Using fixed effects, the model specification unifies the slopes but allows issues to vary in their intercept. The figure, therefore, illustrates the variation in SOU and OR attention and the confidence level of each prediction.

For example, when presidents devote no attention to health issues in their programmatic SOU speech, the predicted attention to this issue in occasional remarks is nearly 3% (2.8%). At the mean attention to this topic at the SOU (5%), presidents devote about a similar (4.9%) share of policy attention in their occasional remarks throughout the following year. At the maximum level of SOU attention (21%), presidents follow up on their programmatic message with 11% average attention in their occasional remarks in the following year. Taken together, the figure illustrates the substantive effect of the presidential attention in the programmatic speech on the voluminous, routine public remarks.

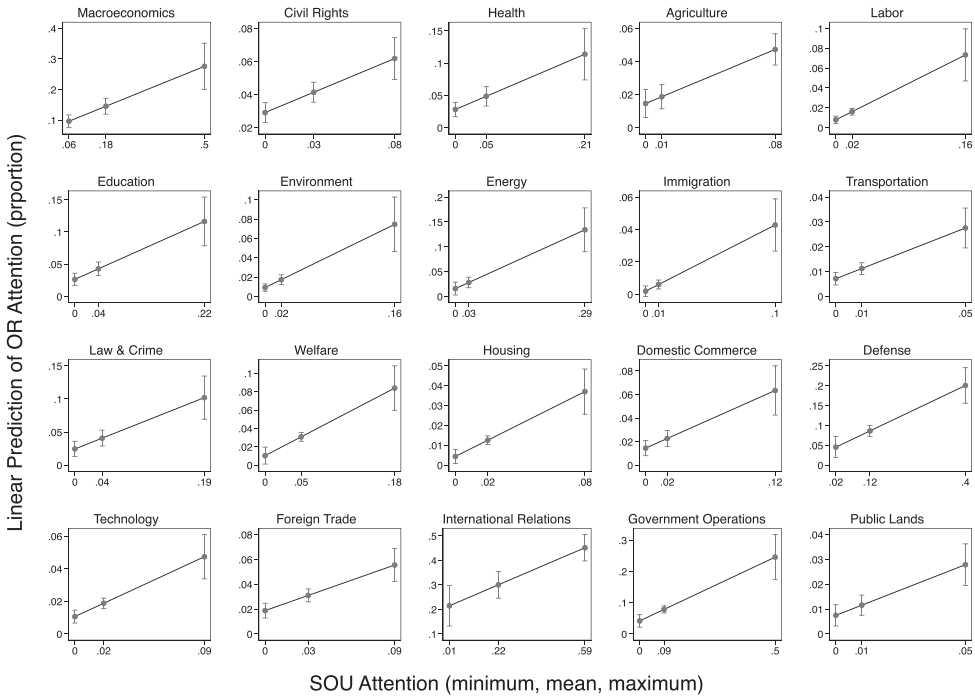

FIGURE 5 Linear prediction of OR attention by topic.

To test the diminishing effect of SOU on OR (H2), we estimated separately the model for each quarter. The quarterly models are identical to the pooled model (Equation 2), except for the now unnecessary control for quarters. The models are summarized in Equation (2). Each model includes a different set of observations—first quarters in a calendar year (January–March), second (April–June), third (July–September), and fourth (October–December).

$$\mathrm{OR}_t = b_1 \mathrm{SOU}_{t,y} + b_2 \mathrm{President}_j + b_4 \mathrm{MajorTopic}_m + \in_j \quad (2)$$

The results are summarized in models 2 through 5 in Table 3. Consistent with Hypothesis 2, the congruence of SOU policy attention and OR attention declines over time. And yet, the positive and significant effect of SOU attention on OR across all quarters renders additional support for Hypothesis 1—the overall congruence of the policy agenda of SOU and OR throughout the year.

Finally, we estimated the effect of events on the correlation between the diversity of policy attention in SOU and the diversity in quarterly attention in OR. Given the limited events data, our models include data from 1953 through 2020. Figure 6 summarizes the diversity in attention in both series.

The two series have similar means (2.16 and 2.21 for OR and SOU, respectively) and variation (standard variation of 0.32 and 0.29, respectively). The plot suggests a weak upward trend in the OR series but not a clear trend in the SOU series. An ARIMA test confirms that the occasional remarks series has a strong AR1 process (0.54, $Z = 12.74$). Both series are stationary (Dickey-Fuller test, Z (SOU) $= -5.518$, Z (OR, 1 lag) $= -5.885$). We, therefore, include a lagged term $(q-1)$ for OR in our model. To test the causal effect of events on diversity in occasional remarks, we included an interaction term between the number of events in each quarter and the diversity in the SOU. We further control for differences between quarters using quarter fixed effects. Equation (3) summarizes model 3. The results are summarized in Table 4.

$$D(\mathrm{OR})_q = b_1 D(\mathrm{SOU})_y + b_2 D(\mathrm{OR})_{q-1} + b_3 \mathrm{Quarter}_q + b_4 \mathrm{Events}_q + b_5 D\left(\mathrm{SOU}_y\right) \times \mathrm{Events}_q + \in \quad (3)$$

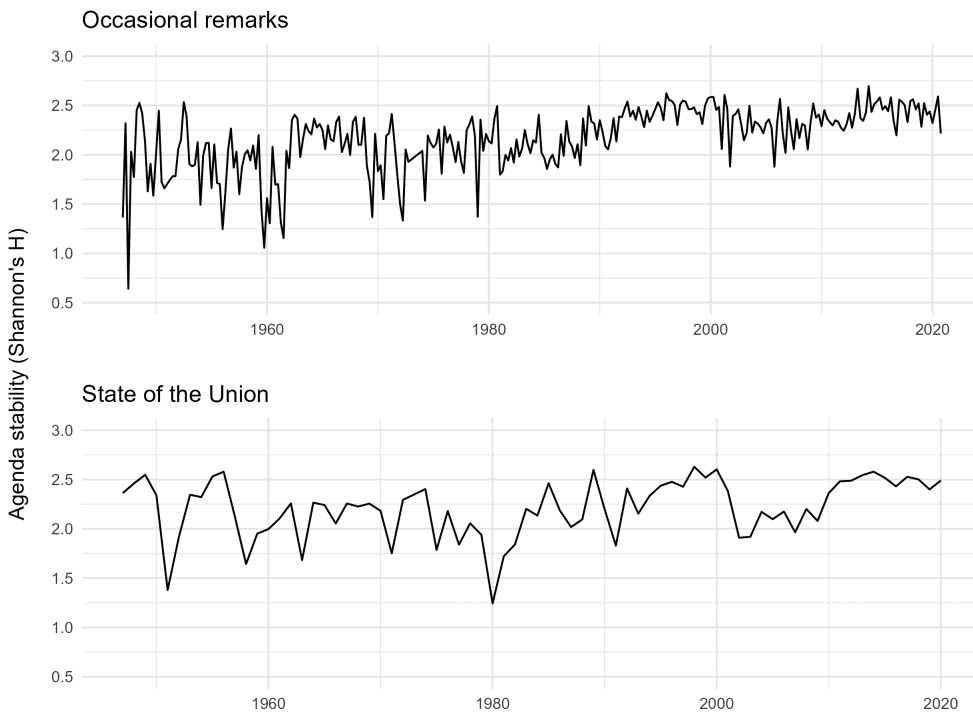

FIGURE 6 Diversity of attention in OR and SOU.

TABLE 4 The effect of policy diversity in SOU on policy diversity in OR.

DV: Policy diversity in OR	(1)	(2)	(3)
Diversity (SOU)	0.120*	0.112*	0.124
	(2.49)	(2.23)	(1.69)
Diversity, OR, $q-1$	0.511***	0.543***	0.543***
	(7.18)	(7.26)	(7.25)
Events (quarter, count)		0.097*	0.224
		(2.19)	(0.54)
Diversity (SOU) * Events (Quarter)			−0.130
			(−0.32)
N	268	268	268
R^2	0.371	0.380	0.380

Note: Standardized beta coefficients; t statistics in parentheses.
*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Model 1 offers a base model to demonstrate the effect of diversity in the State of the Union on diversity of attention in occasional remarks. When presidents address more issues in the State of the Union, we find higher diversity of policy attention in their daily messages. Model 2 adds the events covariate and demonstrates that events *increase* the diversity of policy attention in daily messages. Though the effect is small, it is oppositely signed (compared to *H3a*). Model 3 tests the conditional effect of events on the relationship between the diversity of attention in SOU on the diversity of attention in OR, showing no significant effect. The results show that, at least in the operationalization of events we use here, the number of significant events per quarter does not have the effects we hypothesized (*H3a* and *H3b*).

It is possible that a more refined focus on specific events would reveal a stronger relationship between policy attention in OR and events. For example, our data show that following major mass shooting events, there is an apparent increase in attention to law and crime issues. Following the Sandy Hook shooting on December 14, 2012, attention to law and crime rose 7-fold from a little over 1% of the agenda to nearly 8%. President Obama made it clear that events influence action: "In the coming weeks, I will use whatever power this office holds to engage my fellow citizens—from law enforcement to mental health professionals to parents and educators—in an effort aimed at preventing more tragedies like this." (Obama, December 16, 2012). Action is not only policy making but also a responsibility to educate the American people about the importance of this cause. Similarly, following the Charleston Church shooting on June 15, 2015, attention to law and crime went up from 0% to 10% in the following quarter. Examining this effect more systematically, however, requires a different approach that is beyond the scope of this article.

CONCLUSION

In this article, we built on the concept of dynamic agenda representation to put forth a theory of agenda congruence within U.S. presidential communication. We proposed that policy emphases in major programmatic messages such as State of the Union addresses are followed up by the president's day-to-day communication. We investigated this congruence with a new database of presidential speeches. First, we conducted a deep learning-enhanced sentence-level policy topic coding of various forms of the speeches U.S. presidents made from Truman to Trump. Next, using this database, we desmonstrate that presidents' occasional, day-to-day remarks strongly correlate with the annual policy messages—in this sense, presidents are staying on the democratic script.

The research offers a three-fold contribution to the literature. First, the expansion of the dynamic agenda representation framework to presidential communication yielded a novel theoretical approach with potential relevance beyond presidential studies to other cases of executive politics. The logic of the transmission between programmatic and occasional communication may have wider external validity and can serve as a building block for a more general theory of how policy mandates are translated into executive policymaking and rhetoric.

Second, we applied a deep learning-supported sentence-level policy topic coding for a comprehensive corpus of American presidential speeches. With our fine-tuned RoBERTa language model, we achieved 88.9% accuracy in assigning policy topic codes to sentences. This can be considered a state-of-the-art level performance for the Comparative Agendas Project coding task. Our results show that large language models offer not just a viable but superior modeling strategy vis-á-vis bag-of-words-based supervised models (and when it comes to reliability, they offer superior performance to even double-blind human coding widely considered the gold standard for such tasks).

Third, using these unique data and the innovative methodology to code them, we could test our hypotheses derived from the theoretical discussion related to dynamic agenda representation theory. We investigated how correlated the policy emphases in SOU are to those of the presidents' routine public remarks. We showed that presidents stayed on their democratic script and made efforts to retain the marquee policy topics of their SOU addresses during their more mundane communications. Over time, we see a decline in this congruence as presidents turn to other issues when old issues become less relevant and new circumstances prompt reactions. We argue that this systematic decline offers additional support for the hypothesized role of the State of the Union in setting the rhetorical agenda of the president. Yet, the limited congruence and systematic decline reveal that SOU only partially portray the presidents' rhetorical policy agenda. Future research on the rhetorical presidency need to take a comprehensive approach of the presidents speechmaking efforts in order to assess correctly what presidents say. Such work can inform existing debates about the role and effect of the rhetorical presidency on public opinion, congressional and administrative action, and electoral politics.

Even though our results reinforce the notion that the strategic and routine speechmaking of the American president is far from haphazard when it comes to policy focus, our research design had limitations in multiple respects. Relaxing these limitations could open new avenues for our understanding of the role of presidential communication within the larger context of representative theory.

First, we examined speeches as a closed system of presidential communication and, for the sake of a focused contribution, moved their interplay with campaigns (pledges and platforms), public opinion, and public policy aside. A fuller integration of our research agenda into dynamic representation theory would also investigate these interplays. In traditional mandate theory, the program-to-policy linkage starts with the electoral campaign period, when the policy and rhetorical mandate are formulated and—in many cases—put into a written proposition (manifestos, platforms, planks). A natural extension could look at candidate speeches from big occasions (National Conventions) to county fairs and weigh their policy emphasis, one that also examines ongoing shifts in public support and public agenda. That would create a third pillar for comparing presidential speechmaking from the earliest signs of candidacy to the last days in office.

Second, a deeper look at the functions of various speech types is needed. SOUs are programmatic policy addresses to Congress, while ORs are a leadership tool presidents use to advance their policies and connect with the American people. Presidents may be using these speeches as cheap talk to demonstrate action despite a lack of policy success (or even policy focus). For example, President Trump was known to make repeated claims about immigration, yet he introduced no comprehensive immigration policy proposals during his term starting in 2017. Assessing the relationship between OR and policymaking can better address this concern.

Third, occasional remarks included in our model are defined using the (now somewhat dated) categorization of speeches by the American Presidency Project. Although ORs cover most speeches (75% of all public remarks), they may not sufficiently represent all public events of presidents. Arguably, in other speeches, presidents could be more policy-focused, and these may demonstrate a different policy attention. Additional empirical investigation of other speeches may further test this claim.

Fourth, our data treats all occasional remarks similarly without accounting for exposure or possible impact. Some of these speeches have limited exposure, while others are widely viewed and covered by the news media. Assessing the exposure to every speech (and part of a speech) was beyond the scope of this project. Yet, acknowledging that presidents are aware of the variation in the exposure of their speeches, such a measure can further inform the connection we are suggesting between programmatic messages and day-to-day addresses and situate it within representation theory.

Finally, there is a considerable possibility for extracting additional insights on presidential communication by juxtaposing policy topic emphases and policy position-taking. Our theoretical work and research design are positioned firmly within the saliency and policy agendas tradition as we measured policy emphases as systematic quantitative non-directional assessments of attention. Combining such an analysis with variables related to ideological position-taking and an evaluation of the shifting of such positions over time would marry two hitherto mostly distinctly developed yet major branches of the literature.

ACKNOWLEDGEMENTS

We are thankful for the valuable feedback from the editor and anonymous reviewers as well as the various participants in our presentations of the project at the annual meetings of the APSA, SPSA, the Comparative Agendas Project, COMPTEXT, and the Israeli Political Science Association, the seminar of the Department of Government and Public Policy at CSS in Budapest, and the seminar of the Policy Agendas Project at the University of Texas at Austin. We also acknowledge the help of Kerby Shedden for some of the modeling considerations and of our capable research assistants who did the immense coding required for this project: Hannah Lowenkamp, Csaba Molnár, Ido Rosenberg, and Benjamin Yoel. All errors are ours only.

FUNDING INFORMATION

The research project was supported by the European Union within the framework of the RRF-2.3.1-21-2022-00004 Artificial Intelligence National Laboratory Program; by the EU Horizon 2020 Framework Programme (H2020) under grant agreement no. 101008468 (SLICES); by the Hungarian Academy of Sciences: V-Shift Momentum Project; and by the Israel Science Foundation (ISF no. 1271/18). We are grateful for the possibility to use the HUN-REN Cloud (https://science-cloud.hu/).

DATA AVAILABILITY STATEMENT

Full replication data and scripts are available under the anonymous link: https://figshare.com/s/67b16f90d7ecc8e8cbc4.

ORCID

Miklós Sebők https://orcid.org/0000-0003-0595-2951

ENDNOTES

[1] We define a speech as every public event in which the president gives an oral remark.

[2] Quote taken from Hoffman and Howard (2006, p. 52).

[3] The remaining 25% include major messages to the nation, speeches to a Joint Session of Congress (including SOU), press conferences, inaugural and farewell addresses, speeches at the United Nations, Bill- and Veto-related communications, and radio addresses.

[4] Barack Obama, Remarks at the University of Nebraska Omaha in Omaha, Nebraska Online by Gerhard Peters and John T. Woolley, The American Presidency Project, https://www.presidency.ucsb.edu/node/313386.

[5] For a detailed description of the U.S. codebook see the general codebook available on the CAP website at https://www.comparativeagendas.net/us/datasets. Notably, the comparative CAP codebook contains an additional code (23-Culture) not featured in the U.S. Policy Agendas Project codebook (Eissler & Jones, 2019).

[6] Incoming presidents traditionally do not give a separate address on the State of the Union following their inauguration address. Most of them use their first address to a joint session of Congress (usually delivered in February) as their main programmatic speech. We consider these speeches for Reagan (1981), H.W. Bush (1989), Clinton (1993), W. Bush (2001), Obama

(2009), and Trump (2017). President Eisenhower did not deliver an oral State of the Union address in 1956 and therefore we used the written address instead. President Nixon submitted six separate written addresses in 1973. We did not include them in our data.

[7] The categorization of the occasional remarks in the American Presidency Project has changed since we downloaded and coded the data. We present here our analysis based on the original categorization scheme.

[8] For the fine-tuning of the model, we used a batch size of 32, learning rate of 3e−5 and a dropout rate in the final classification layer of 0.1 to avoid overfitting. Due to hardware constraints, the maximum sentence length that the RoBERTa model can process with these hyperparameters is 120 words. Sentences longer than 120 words were automatically truncated to 120 and then coded. Our data include 846 such sentences (0.08% of the data).

[9] By combining the Defense and International Affairs topics, the difference between expert and machine coded label proportions almost halved.

[10] This can be formally written as $\text{Stability} = 1 - \frac{\sum_1^q \text{abs}\left(\text{OR}_q - \text{SOU}_{yq}\right)}{2}$.

REFERENCES

Barabas, Jason. 2008. "Presidential Policy Initiatives: How the Public Learns about State of the Union Proposals from the Mass Media." *Presidential Studies Quarterly* 38(2): 195–222.

Barrett, Andrew W. 2004. "Gone Public: The Impact of Going Public on Presidential Legislative Success." *American Politics Research* 32(3): 338–70.

Baum, Matthew A., and Samuel Kernell. 1999. "Has Cable Ended the Golden Age of Presidential Television?" *American Political Science Review* 93(1): 99–114.

Baumgartner, Frank R., Christian Breunig, and Emiliano Grossman. 2019. "The Comparative Agendas Project: Intellectual Roots and Current Developments." In *Comparative Policy Agendas*, edited by Frank R. Baumgartner, Christian Breunig, and Emiliano Grossman, Ch. 1, 3–16. Oxford: Oxford University Press.

Bevan, Shaun. 2019. "Gone Fishing: The Creation of the Comparative Agendas Project Master Codebook." In *Comparative Policy Agendas: Theory, Tools, Data*, edited by Frank R. Baumgartner, Christian Breunig, and Emiliano Grossman, 17–34. Oxford: Oxford University Press.

Bevan, Shaun, and Will Jennings. 2014. "Representation, Agendas and Institutions." *European Journal of Political Research* 53(1): 37–56.

Bevan, Shaun, and Anne Rasmussen. 2020. "When Does Government Listen to the Public? Voluntary Associations and Dynamic Agenda Representation in the United States." *Policy Studies Journal* 48(1): 111–32.

Blumenthal, Sidney. 1980. *The Permanent Campaign: Inside the World of Elite Political Operatives*. Boston, MA: Beacon Press.

Boydstun, Amber E. 2013. *Making the News: Politics, the Media, and Agenda Setting*, Chicago, London: The University of Chicago Press.

Budge, Ian, and Richard I. Hofferbert. 1990. "Mandates and Political Outputs: U.S. Party Platforms and Federal Expenditures." *American Political Science Review* 84(1): 111–31.

Campbell, Karlyn Kohrs, and Kathleen Hall Jamieson. 2008. *Deeds Done in Words: Presidential Rhetoric and the Genres of Governance*. Chicago, IL: University of Chicago Press.

Canes-Wrone, Brandice. 2006. *Who Leads Whom?: Presidents, Policy, and the Public*. Chicago, IL: University of Chicago Press.

Cavari, Amnon. 2013. "The Short-Term Effect of Going Public." *Political Research Quarterly* 66(2): 336–51.

Cavari, Amnon. 2017. *The Party Politics of Presidential Rhetoric*. New York: Cambridge University Press.

Cavari, Amnon, Benjamin Yoel, and Hannah Lowenkamp. 2021. "In the Name of the President." *American Politics Research* 49: 666–80.

Cavari, Amnon, Maoz Rosenthal, and Ilana Shpaizman. 2023. "Reevaluating the Policy Success of Private Members Bills". *Research & Politics* 10(2). https://doi.org/10.1177/20531680231181750

Cohen, Jeffrey E. 1995. "Presidential Rhetoric and the Public Agenda." *American Journal of Political Science* 39(1): 87–107.

Cohen, Jeffrey E. 2008. *The Presidency in the Era of 24-Hour News*. Princeton, NJ: Princeton University Press.

Cohen, Jeffrey E. 2010. *Going Local: Presidential Leadership in the Post-Broadcast Age*. New York, NY: Cambridge University Press.

Cohen, Jeffrey E. 2012. *The President's Legislative Policy Agenda, 1789–2002*. New York, NY: Cambridge University Press.

Cummins, Jeff. 2008. "State of the Union Addresses and Presidential Position Taking: Do Presidents Back their Rhetoric in the Legislative Arena?" *The Social Science Journal* 45(3): 365–81.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. "Bert: Pre-Training of Deep Bidirectional Transformers for Language Understanding." *arXiv preprint arXiv:1810.04805*.

Edwards, George C. 2003. *On Deaf Ears: The Limits of the Bully Pulpit*. New Haven, CT: Yale University Press.

Edwards, George C. 2009. *The Strategic President: Persuasion and Opportunity in Presidential Leadership*. New Haven, CT: Princeton University Press.

Edwards, George C. 2012. *Overreach: Leadership in the Obama Presidency*. Princeton, NJ: Princeton University Press.

Edwards, George C. 2016. *Predicting the Presidency: The Potential of Persuasive Leadership*. Princeton, NJ: Princeton University Press.

Edwards, George C. 2021. *Changing their Minds? Donald Trump and Presidential Leadership*. Chicago, IL: University of Chicago Press.

Eissler, Rebecca, and Bryan D. Jones. 2019. "The U.S. Policy Agendas Project." In *Comparative Policy Agendas: Theory, Tools, Data*, edited by Frank R. Baumgartner, Christian Breunig, and Emiliano Grossman, 184–90.

Eshbaugh-Soha, Matthew. 2016. *Going Public and Presidential Leadership*. Oxford: Oxford Research Encyclopedia of Politics.

Eshbaugh-Soha, Matthew, and Eric Gonzalez Juenke. 2022. "The Politics of the President's Immigration Rhetoric." *American Politics Research* 50(1): 117–30.

Eshbaugh-Soha, Matthew, and Thomas Miles. 2011. "Presidential Speeches and the Stages of the Legislative Process." *Congress and the Presidency* 38(3): 301–21.

Eshbaugh-Soha, Matthew, and Jeffrey S. Peake. 2004. "Presidential Influence over the Systemic Agenda." *Congress and the Presidency* 31(2): 181–201.

Eshbaugh-Soha, Matthew, and Jeffrey S. Peake. 2005. "Presidents and the Economic Agenda." *Political Research Quarterly* 58(1): 127–38.

Fett, Patrick J. 1992. "Truth in Advertising: The Revelation of Presidential Legislative Priorities." *Western Political Quarterly* 45(4): 895–920.

Fett, Patrick J. 1994. "Presidential Legislative Priorities and Legislators' Voting Decisions: An Exploratory Analysis." *The Journal of Politics* 56(2): 502–12.

Fontaine, Sam A., and Daniel M. Gomez. 2020. "Going Social: A Comparative Analysis of Presidents' Official and Social Media Messages." *Presidential Studies Quarterly* 50(3): 507–38.

Gergen, David. 2008. *Eyewitness to Power the Essence of Leadership Nixon to Clinton*. New York: Simon & Schuster.

Haeder, Simon F., and Jacqueline Chattopadhyay. 2022. "The Power of a Tweet? Social Media, Presidential Communication, and the Politics of Health." *Presidential Studies Quarterly* 52(2): 436–73.

Heith, Diane J. 2013. *The Presidential Road Show: Public Leadership in an Era of Party Polarization and Media Fragmentation*. Boulder, CO: Paradigm Publishers.

Heith, Diane J. 2020. *The End of the Rhetorical Presidency? Presidential Leadership in the Trump Era*. New York: Routledge.

Hoffman, Donna R., and Alison D. Howard. 2006. *Addressing the State of the Union: The Evolution and Impact of the President's Big Speech*. Boulder, CO: Lynne Rienner Publishers.

Jennings, Will, Shaun Bevan, and Peter John. 2011. "The Agenda of British Government: The Speech from the Throne, 1911-2008." *Political Studies* 59(1): 74–98.

Jennings, Will, Shaun Bevan, Arco Timmermans, Gerard Breeman, Sylvain Brouard, Laura Chaqués-Bonafont, Christoffer Green-Pedersen, Peter John, Peter B. Mortensen, and Anna M. Palau. 2011. "Effects of the Core Functions of Government on the Diversity of Executive Agendas." *Comparative Political Studies* 44(8): 1001–30.

Jennings, Will, and Peter John. 2009. "The Dynamics of Political Attention: Public Opinion and the Queen's Speech in the United Kingdom." *American Journal of Political Science* 53(4): 838–54.

Jones, Bryan D., Frank R. Baumgartner, Sean M. Theriault, Derek A. Epp, Cheyenne Lee, and Miranda E. Sullivan. 2023. *Policy Agendas Project: Codebook*. https://www.comparativeagendas.net/us

Kernell, Samuel. 2007. *Going Public: New Strategies of Presidential Leadership*, 4th ed. Washington, DC: CQ Press.

Körösényi, András, and Miklós Sebők. 2018. "From Pledge-Fulfilment to Mandate-Fulfilment." *Intersections* 4: 115–32.

Kumar, Martha Joynt. 2007. *Managing the President's Message: The White House Communications Operation*. Baltimore, MD: The Johns Hopkins University Press.

Light, Paul. 1999. *The President's Agenda: Domestic Policy Choice from Kennedy to Clinton*. Baltimore, MD: Johns Hopkins University Press.

Liu, Yinhan, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. "RoBERTa: A Robustly Optimized BERT Pretraining Approach." *arXiv preprint arXiv:1907.11692*.

Loftis, Matthew W., and Peter B. Mortensen. 2020. "Collaborating with the Machines: A Hybrid Method for Classifying Policy Documents." *Policy Studies Journal* 48(1): 184–206.

Lovett, John, Shaun Bevan, and Frank R. Baumgartner. 2015. "Popular Presidents Can Affect Congressional Attention, for a Little While: Presidents Can Affect Congressional Attention." *Policy Studies Journal* 43(1): 22–43.

Manin, Bernard, Adam Przeworski, and Susan C. Stokes. 1999. "Elections and Representation." In *Democracy, Accountability, and Representation*, edited by Adam Przeworski, Susan C. Stokes, and Bernard Manin. New York, NY: Cambridge University Press.

Mansbridge, Jane. 2003. "Rethinking Representation." *American Political Science Review* 97(4): 515–28.

Mortensen, Peter Bjerre, Christoffer Green-Pedersen, Gerard Breeman, Laura Chaqués-Bonafont, Will Jennings, Peter John, Anna M. Palau, and Arco Timmermans. 2011. "Comparing Government Agendas: Executive Speeches in The Netherlands, United Kingdom, and Denmark." *Comparative Political Studies* 44(8): 973–1000.

Newman, Brian, and Andrew Forcehimes. 2010. "'Rally Round the Flag' Events for Presidential Approval Research." *Electoral Studies* 29(1): 144–54.

O'Brien, Shannon Bow. 2019. *Why Presidential Speech Locations Matter: Analyzing Speechmaking from Truman to Obama*. New York: Palgrave Macmillan.
Ornstein, Norman J., and Thomas E. Mann. 2000. *The Permanent Campaign and its Future*. Washington, DC: American Enterprise Institute; Brookings Institution.
Ostrom, Charles, Alon Kraitzman, and Brian Newman. 2021. "Explaining Presidential Approval, 1993–2020: An Examination of the Clinton, Bush, Obama and Trump Presidencies." Advance. https://doi.org/10.31124/advance.16575062.v1
Ouyang, Yu, and Richard W. Waterman. 2020. "Trump, Twitter, and the American Democracy." In *Trump, Twitter, and the American Democracy* 131–61. Cham: Springer International Publishing.
Peake, Jeffrey S. 2001. "Presidential Agenda Setting in Foreign Policy." *Political Research Quarterly* 54(1): 69–86.
Rottinghaus, Brandon. 2010. *The Provisional Pulpit: Modern Presidential Leadership of Public Opinion*. College Station, TX: Texas A&M University Press.
Royed, Terry J. 1996. "Testing the Mandate Model in Britain and the United States: Evidence from the Reagan and Thatcher Eras." *British Journal of Political Science* 26(1): 45–80.
Russell, Annelise, and Rebecca Eissler. 2022. "Conditional Presidential Priorities: Audience-Driven Agenda Setting." *American Politics Research* 50(4): 545–9.
Sebők, Miklós, and Zoltán Kacsuk. 2021. "The Multiclass Classification of Newspaper Articles with Machine Learning: The Hybrid Binary Snowball Approach." *Political Analysis* 29(2): 236–49.
Sebők, Miklós, Zoltán Kacsuk, and Ákos Máté. 2022. "The (Real) Need for a Human Touch: Testing a Human–Machine Hybrid Topic Classification Workflow on a New York Times Corpus." *Quality & Quantity* 56(5): 3621–43.
Thomson, Robert. 2001. "The Programme to Policy Linkage: The Fulfilment of Election Pledges on Socio-Economic Policy in The Netherlands, 1986–1998." *European Journal of Political Research* 40(2): 171–97.
Vaswani, Ashish. 2017. "Attention is All You Need." *Advances in Neural Information Processing Systems* 30: I.
Villalobos, José D., Justin S. Vaughn, and Julia R. Azari. 2012. "Politics or Policy? How Rhetoric Matters to Presidential Leadership of Congress." *Presidential Studies Quarterly* 42(3): 549–76.
Wankmüller, Sandra. 2022. "Introduction to Neural Transfer Learning with Transformers for Social Science Text Analysis." *Sociological Methods & Research*. https://doi.org/10.1177/00491241221134527
Whitford, Andrew B., and Jeff Yates. 2009. *Presidential Rhetoric and the Public Agenda: Constructing the War on Drugs*. Baltimore, MD: Johns Hopkins University Press.
Wood, B. Dan. 2007. *The Politics of Economic Leadership: The Causes and Consequences of Presidential Rhetoric*. Princeton, NJ: Princeton University Press.
Wood, B. Dan, and Jeffrey S. Peake. 1998. "The Dynamics of Foreign Policy Agenda Setting." *American Political Science Review* 92(1): 173–84.
Yates, Jeff, and Andrew Whitford. 2005. "Institutional Foundations of the President's Issue Agenda." *Political Research Quarterly* 58(4): 577–85.

AUTHOR BIOGRAPHIES

Amnon Cavari is Associate Professor and head of the Institute for Liberty and Responsibility at the Lauder School of Government, Diplomacy and Strategy at Reichman University, Israel. Prof. Cavari's main research interests are in the interrelationship between actions of elected officials and public opinion in the United States and in Israel. He is the author of *The Party Politics of Presidential Rhetoric*.

Akos Mate is a computational social scientists whose research interest are political economics and quantitative methodology. He is a research fellow at the Centre for Social Sciences, Budapest. He also teaches as visiting faculty at the Central European University, Vienna and served as a consultant for the IMF's Independent Evaluation Office.

Miklós Sebők is a Senior Research Fellow at the HUN-REN Centre for Social Sciences (CSS), Budapest. He serves as the research co-director of the Artificial Intelligence National Lab at CSS, the principal investigator of the V-SHIFT Momentum research project, and the convenor of the COMPTEXT conference. His main research interest lies at the intersection of policy studies and natural language processing.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Cavari, Amnon, Akos Mate and Miklós Sebők. 2024. "Staying on the Democratic Script? A Deep Learning Analysis of the Speechmaking of U.S. Presidents." *Policy Studies Journal* 00 (0): 1–21. https://doi.org/10.1111/psj.12534.