Working Paper Series

SoDa LABS

Social [science] insights from alternative data

Est. 2018

Tracking Policy-relevant Narratives of Democratic Resilience at Scale: from experts and machines, to AI & the transformer revolution

Simon D Angus

SoDa Laboratories Working Paper Series

No. 2024-07

REF

Simon D Angus (2024), SoDa Laboratories Working Paper Series No. 2024-07, Monash Business School, available at http://soda-wps.s3-website-ap-southeast-2.amazonaws.com/RePEc/ajr/sodwps/2024-07.pdf

PUBLISHED ONLINE

10 December 2024

SoDa Laboratories

https://www.monash.edu/business/soda-labs/

MONASH University | MONASH BUSINESS SCHOOL

AACSB ACCREDITED EFMD EQUIS ACCREDITED AMBA ACCREDITED

ABN 12 377 614 012 CRICOS Provider Number: 00008C

XXX (2024), 1–21
doi:10.1017/xxxx

PRE-PRINT

Tracking Policy-relevant Narratives of Democratic Resilience at Scale: from experts and machines, to AI & the transformer revolution

Simon D. Angus[1]

[1]Dept. of Economics, Monash University, Clayton, 3800, VIC, Australia. [1]SoDa Laboratories, Monash Business School, Clayton, 3800, VIC, Australia. .
Corresponding author. E-mail: simon.angus@monash.edu

Received xx xxx xxxx

Keywords: Computational linguistics; Political discourse analysis; Natural Language Processing; Quantitative Social Science; AI in policy research

Abstract
Democratic resilience is as much about the narratives of our nation we affirm, as the institutions that enshrine our values and laws, a fact re-affirmed by scholarship across many branches of social science in recent decades. For policymakers and quantitative social scientists, analysing or tracking public discourse through the lens of narrative and framing has historically involved the annotation of texts by hand, placing severe limitations on the scale and modality of discourse under inquiry. In this study, we consider a variety of tools from the field of computational linguistics, which either automate the standard approach to textual annotation, or introduce entirely new ways of conceptualising 'text as data', opening up new horizons for the tracking of public narratives of democratic resilience. In particular, we assess the regime-shift occurring in natural language processing and artificial intelligence brought about by the advent of the transformer architecture. These new tools offer, perhaps for the first time, the 'holy grail' of the quantitative social scientist: the ability to identify, accurately, and efficiently, nuanced narratives in text at scale. We conclude by contributing data and research recommendations for public stakeholders who wish to see these opportunities realised.

Policy Significance Statement

Good policy starts with good information. In the case of policy relating to democratic resilience, whilst there is a very large set of choices available to the policymaker who desires to strengthen democracy, there is a remarkably slender set of rich information on citizen narratives of democracy on which to draw. Whilst surveys will always be important, we are now in an age where voluminous alternative data forms are arising in social media, digital media, and political discourse in real time. These new sources present a huge opportunities for policymakers to identify, track, and analyse rich belief structures – how individuals frame issues and events – in near real time. In this paper, we ask whether advances in natural-language-processing (NLP) and in particular, recent breakthroughs in generative AI technologies, can be applied to democratic narrative analysis at scale. We further identify the centrality of harmonised data sources to any future endeavour, and make recommendations for policymakers to foster these methods and tools.

1. Introduction

Story-telling has always been a powerful form of human expression: we venerate our literary giants; and we encode our culture, history and identity in the stories that we tell our children and each other in so many private moments. Yet for democracies, story telling, or more formally ‘narratives’, seem to take on a particularly potent guise. Each election cycle, those of us fortunate enough to live in an open, democratic society, will encounter political leaders who’s chief form of political speech will be in narrative form: a vision for a better future; a depiction of an opponent’s ‘world’; a personal life ‘story’, connecting the leader to the lounge-room.

Policy experts, economists, sociologists and political scientists have long theorised over the role of public narratives in our national political and policymaking journey. For some, narratives have the power to both initiate and catalyse major economic events, from the Great Depression to financial crises (Shiller, 2019), whilst others articulate the common finding that public narratives are at their most powerful when they resonate with personal convictions *already held* (Polletta and Callahan, 2017). Indeed, reflecting on the potency of Trump’s campaign tactics, Polletta and Callahan note that it was the ‘small stories’ heard through a complex web of radio, TV, social media, and personal networks which (p.13), ‘meshed with the experiences of others in a way that made them all seem personal.’ More forceful of all, narratives have been rediscovered by some as providing a kind of ontology of self – public narrative as ‘identity compass’ to the inner self. According to Somers (1994) (p.606) (emphasis added),

> It matters not whether we are social scientists or subjects of historical research, but that all of us come to **be** who we **are** (however ephemeral, multiple, and changing) by being located or locating ourselves (usually unconsciously) in social narratives **rarely of our own making**.

In other words, we now recognise that public narratives matter, not only for how public debate proceeds, but also, for the composition and sense-making of individuals themselves.

Yet if the power of public narratives to shape individuals and society was already recognised in the 90s, how much more should we pay attention to them today, given the unprecedented acceleration in the availability, volume and velocity (Mauro et al., 2016) of public discourse surging through our democracies? Parallel mega-trends including the ubiquity of the internet, social media platforms, the 24/7 news cycle, the spectre of mis-information, ‘echo-chambers’ and geo-political interference; each contribute to a sense that for those who carry the fire for our democracies, characterising, mapping, and tracking the maelstrom of competing narratives has become an *essential* task.

But are our methodological tools up to the task? Are there accepted standards for analysing public discourse to identify and track key narratives of democratic threat and resilience?

In this paper, we shall take stock of current and emerging capabilities to track narratives related to democratic resilience at scale, with the overall purpose of measurement, diagnosis and assessment of how public narratives shape, and are influenced by, social connection and democratic resilience. In so doing, we shall encounter the enormous potential inherent in the latest wave of language modelling technology, opening up new research frontiers to harness this potential in ways that explicitly seek to strengthen democracy. Given that we see few, if any examples of this research to date, we conclude by proposing a new computational research agenda that takes advantage of these tools for strengthening democracy.

2. Narratives, Issue Framing, & Democratic Resilience

2.1. Narratives & Framing: a two-way street

A common approach to classifying narratives for quantitative social science, is to draw on theories of *issue-framing* (or just ‘framing’). Here, one or more positions on any given topic or issue can be identified by examining how a communicator crafts their message on the topic, i.e. how they ‘frame’ the topic. Entman (1993) defines framing thus (p.52) (emphasis added),

> To frame is to **select** some aspects of a perceived reality and make them **more salient** in a communicating text, in such a way as to **promote** a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation for the item described.

Key in Entman’s definition is *selection* and *saliency*, to serve the objective of *promotion* of the speaker’s particular messaging on the topic. As such, issue framing can be seen as a specific aspect of narrative. Where narrative involves characters, events, and often a temporal component, framing is concerned with the specific treatment of an issue, event, or topic by the speaker who wishes to convince the receiver of the same. Chong and Druckman (2007) makes this connection in their more recent theorising (p.104) (emphasis added),

> The major premise of framing theory is that an issue can be viewed from a variety of perspectives and be construed as having implications for multiple values or considerations. Framing refers to the process by which people develop a **particular conceptualisation** of an issue or **reorient their thinking** about an issue.

Critically then, issue-framing applies equally to the speaker *and* receiver. The speaker seeks to promote a particular framing of an issue (*ala* Entman), with the objective of bringing the receiver to the same perspective (*ala* Chong & Druckman).

For example, Benier et al. (2020)’s study of ‘African gangs’ in Australia fits well into these definitions. On the one hand, they document the way that the then Federal Minister for Immigration and Citizenship positioned the death of a young South Sudanese man as evidence of the failure of the South Sudanese community to integrate into Australian society, sparking a framing of ‘African gangs’ that would be re-ignited in the media for years to come. Whilst on the other, in a survey study, they show how individual prejudice towards Africans and a respondent’s political preference, correlate strongly with perceptions of local crime levels. In other words, they find evidence for the way that public framing of Africans in Melbourne was then reflected in personal framing of the same.

But framing is not merely an individual consideration, in democracies, individuals freely join with others to form *groups*, groups join again to form *movements*. And here, at these higher orders of organisation, social theorists see that framing applies equally well. Indeed, there is an important *interplay* between the framing of the group or movement and the individuals who compose it. Snow and Benford (1988) write (p.198),

> Movements function as carriers and transmitters of mobilizing beliefs and ideas, to be sure; but they are also actively engaged in the production of meaning for participants, antagonists, and observers. ... They frame, or assign meaning to and interpret, relevant events and conditions in ways that are intended to mobilize potential adherents and constituents, to garner bystander support, and to demobilize antagonists.

Furthermore, Snow & Bedford go on (p.199) to decompose framing in social movements into three ‘core’ tasks: **diagnosis** “of some event or aspect of social life as problematic and in need of alteration”; **prognostic** “a proposed solution to the diagnosed problem that specifies what needs to be done”; and **motivational** “a call to arms or rationale for engaging in ameliorative or corrective action”. Indeed, they argue that that it is the degree to which a social movement is able to achieve each of these three tasks that will determine the movement’s success or failure. In other words, for social movements, successful framing is everything.

2.2. Narratives & Democratic Resilience

If framing is critical to the success of movements, and narratives have the kind of power to shift economic-systems for better or worse (Shiller, 2019), then it follows that successful democracies, and so, policymakers who carry the fire for their resilience, will encourage the healthy competition of many social movements, and so, their narratives, towards prosperity and progress for all.

Yet there are inherent challenges in getting this balance right. Indeed, according to Holloway and Manwaring's definition of democratic resilience, a population riven by discordant framings is a threat to democracy itself. For them, democratic resilience is that quality of a democratic system which can stand despite the narrowest intersection of agreement on the issues that matter. Namely, 'democratic resilience' is (p.76),

> A system's capacity to withstand a major shock such as the onset of extreme polarization and to continue to perform the basic functions of democratic governance – electoral accountability, representation, effective restraints on excessive or concentrated power, and collective decision-making.

As such, and integrating Snow and Benford's decomposition of framing in social movements, a resilient democracy is one which can hold on to its foundational apparatus, even when the society at large is captured by highly opposing diagnoses, prognoses, or motivations regarding the issues of the day. And, as mentioned earlier, Holloway and Manwaring (2023)'s wide-lens review of literature on democratic resilience finds that (Table 2, p.72) 'diversity (of groups, individuals)', 'learning', and 'community participation and inclusion' are amongst the common characteristics of resilient systems. So democratic resilience is about finding ways for diverse narratives to *co-exist* – i.e. to challenge the status-quo, and/or seek inclusion – whilst at the same time, being wary of narratives which foment excessive polarisation, distrust of electoral institutions and representation, and undermine social cohesion. The most resilient democracy may be able to withstand the worst of these narratives, but recent international experience suggests that democratic systems can be surprisingly susceptible to hitherto marginal narratives.

The difficulty, of course, is identifying which narratives are 'healthy' and which are 'pernicious' to the democratic order. Here, the literature offers up a number of recognisable narratives that any democratic society may wish to be wary of. We briefly consider three.

First, the **'Us vs. Them'** narrative is the most transparently polarising, reducing complex, multi-dimensional issues down to a single dimension, and at its worst, encouraging a kind of 'purity test' of belief for those who adhere to one 'side' of the political landscape or the other (McCoy et al., 2018; McCoy and Somer, 2019; Hetherington and Rudolph, 2015; Iyengar et al., 2019). Despite wide variation in how a specific country becomes extremely polarised, authors contend that the subsequent undermining of democracy follows a very common path: political opponents dis-engage from meaningful communication and compromise; government reform agendas are gridlocked; each side dismantles the other's legislative agenda immediately on gaining power; and in the end-state, democratic transfer of power falls under restrictions of civil liberties and sees the military's entry to replace government.

Relatedly, narratives of nationalism can also work to undermine a plural, inclusive democratic society. As argued by Mylonas and Tudor (2021), we should distinguish between nationalism which draws on ideals of freedom and justice (e.g. the American and French Revolutions), and (p.111), "**exclusionary nationalism** (also called ethnic or essentialist nationalism)". Exclusionary nationalism shares rhetorical approaches with the 'us vs. them' narrative in its reductionism, i.e. driving a hierarchical narrative along racial, ethnic or religious grounds, 'othering' some group or groups in society to achieve some idealised, patriotic objective.

Third, as Algan et al. (2017) ominously begin, 'a spectre is haunting Europe and the West – the spectre of **populism**'. Populism is perhaps the narrative which is most directly opposed to the democratic project, for its proponents set themselves against the very institutions of democracy itself – established political parties, established institutions of state, even the judiciary, are all said to be held captive by 'elites', set against the best interests of the common people. Here, former president Donald Trump's dark framing of modern America, promising to 'drain the swamp' (i.e. of elites), and 'make America great again', spoke to the felt experiences of the American people and dove-tailed with right-wing programming across print, radio, TV and online (Polletta and Callahan, 2017; Norris and Inglehart, 2019).

The message was particularly appealing to those who felt like ‘outsiders’ in their own country due to social and economic marginalisation (Gidron and Hall, 2020).

But how prevalent are these narratives in a modern democracy? Are narratives of populism or exclusionary nationalism already making themselves heard in the United Kingdom, Australia, or the United States? Has there been a change over time? Where? Is this only for the online ‘echo-chambers’ or do we see growing evidence of these narratives in our public discourses arising in Parliament, print or on the air-waves?

The answer to all these questions requires empirical tools.

2.3. Tracking Public Narratives, Social Listening & Surveillance

Before moving to empirical methods, it is worth spending a moment to distinguish present considerations from ‘surveillance’ orientated methods, or what is sometimes referred to as ‘surreptitious social listening’. Listening – the active or passive act of receiving/collecting information from people in a given context – can take a variety of forms, best considered as sitting on a spectrum (Martino, 2020). According to such typologies, tracking of public narratives related to social cohesion and democratic resilience, sits in the ‘active listening’ category, and, according to Martino seeks to *enhance* democratic and foreign policy objectives by promoting trust and understanding. In a word, the government or actor who actively listens to citizens and counterparts, including those who hold opposing views, demonstrates a long-term commitment to dialogue, relation-building, and engagement. Importantly, such active listening, or ‘public discourse tracking‘ as we consider here, is far away from *surreptitious* listening, or what is otherwise known as surveillance or spying. Here, there are spaces for legitimate uses of such methods in democracies, but such uses must carefully balance personal rights and freedoms with policing and counter-terrorism objectives (Hagen and Lysne, 2016; Moore, 2011). Importantly, it should be acknowledge that all surveillance carries harm (Richards, 2013) either because of a chilling effect on civil liberties (Penney, 2016), or by introducing unhealthy power dynamics due to information asymmetries. Nevertheless, tracking public narratives related to democracy, where it seeks to understand, characterise, build trust, and engage, aims to support democratic values, not undermine them, and should be kept distinct from states or actors who seek to infiltrate the private sphere to advance coercive, controlling, or chilling purposes.

3. Narrative Tracking in Public Discourse

Public narratives, and the framings they carry, have most commonly been tracked through longitudinal opinion surveys, or via the (human) coding (or ‘labelling’[1]) of political, news or other texts via standard approaches. With the advent of more advanced natural-language-processing (NLP) tools over the past two decades, and in particular, with the rise of powerful large-language-model (LLM) generative artificial intelligence (AI) systems, there has been a rapid growth in methodologies which seek to automate labelling tasks, at scale, with near human levels of accuracy.

The ultimate aim of such efforts is to generate a structured dataset where each record represents a given *text*, *meta-data* or *attributes* related to that text (e.g. the date, speaker, party affiliation, location, medium etc.), and a series of *labels* which indicate whether a given framing relating to some issue exists or does not exist in that text. From there, the wide array of standard empirical tools can be applied to generate analytical objects (e.g. time-series, histograms), statistical tests, and causal inferences.

We focus here on non-survey methods, charting the pathway from labor-intensive labelling to automated approaches.[2]

3.1. Coding frames in discourse: the standard approach

In their review of framing theory, Chong and Druckman (2007) note that the identification of frames in communication by domain experts has become a, 'virtual cottage industry' (p.106). And, whilst they note that no uniform definition of how to proceed exists, the 'most compelling studies' follow six standard steps, outlined below.

1. **Identify the issue** – Chong and Druckman note that frames only relate to a *specific* issue or event. As such, the necessary starting point of any framing analysis is to identify the issue or event. For example, the 'Indigenous Voice to Parliament' (issue), or 'Climate Change' (issue), or the 'Moomba Riots' (event).
2. **Isolate an attitude (dimension)** – Most issues will have multiple attitudes, or dimensions, on which frames may exist. For example, the rise of populism in western democracies could be considered from the perspective of trust in government, political representation, challenges of integration, or a series of other considerations. Multiple frames may arise in connection with each of these attitudes.
3. **Define frames** – Given an issue, and a dimension of analysis of that issue, one must then identify the frames to study. Typically, this is an inductive process. For example, the issue of *immigration*, could be studied in terms of *attitudes* towards refugees, through frames such as 'humanitarian approach' versus 'border security approach'.
4. **Select sources** – Next, source materials are identified and any qualifying filters applied. For example, Australian Parliamentary Hansard, online news websites, social media, op-ed pieces, transcripts of radio shows or other online media. An *inclusion criteria* is typically required to select only those texts which relate specifically to the *issue* or *event* at hand, and speak to the *attitude* or *dimension* in question. Other considerations include the length of the text[3], the language, and quality of the text.
5. **Develop validated coding scheme** – Perhaps the most critical step is then undertaken: the development of a validated method to code texts for the presence of one or other frame earlier defined. Manual approaches to this step typically see the coders (those who will do the coding) studying representative samples of the source text to examine the language used to represent a given frame, and so, inductively determine positive and negative examples of the existence or absence of a frame in a sample of text. Automated coding approaches either use human labelled samples to *train* the automated system in the identification of the frames, or, human generated vocabulary or phrases which coders would traditionally use, and which can be used as inputs to automated labelling systems. Either way (human or automated), the coding method must be *validated* to ensure it has high accuracy (precision), and ideally, identifies close to all of the texts which align with the frames in the study, missing only a small fraction, if at all (recall). For human coding, the method should also be assessed for *reliability*, that is, the extent to which multiple coders provide consistent (within, and between) coding of the same texts[4].
6. **Undertake coding** – Finally, the coding method is applied to the qualifying texts identified earlier.

Examples of the standard approach in practice include Terkildsen and Schnell (1997) who analyse weekly news coverage of the women's movement over four decades for the presence of frames related to sex roles, feminism/anti-feminism, political rights, and economic rights, and Jennings and John (2009) who code discourse frames including social inclusion/exclusion in successive releases of the poverty reduction strategy in Ontario, Canada. The approach is also used at the level above frames, i.e. to identify the presence or absence of a given set of issues in texts such as in Smith-Carrier and Lawlor (2016) who code the Queen's speech to Parliament for policy areas and join this to Gallup surveys on 'the most urgent problem' to examine whether the government agenda responds to public opinion.

3.2. 'Traditional' NLP approaches to automation

Whilst the standard approach is still widely used and accepted in rigorous analysis of public discourse, the recent digitisation of all areas of human social, economic and political life, and the massive quantity

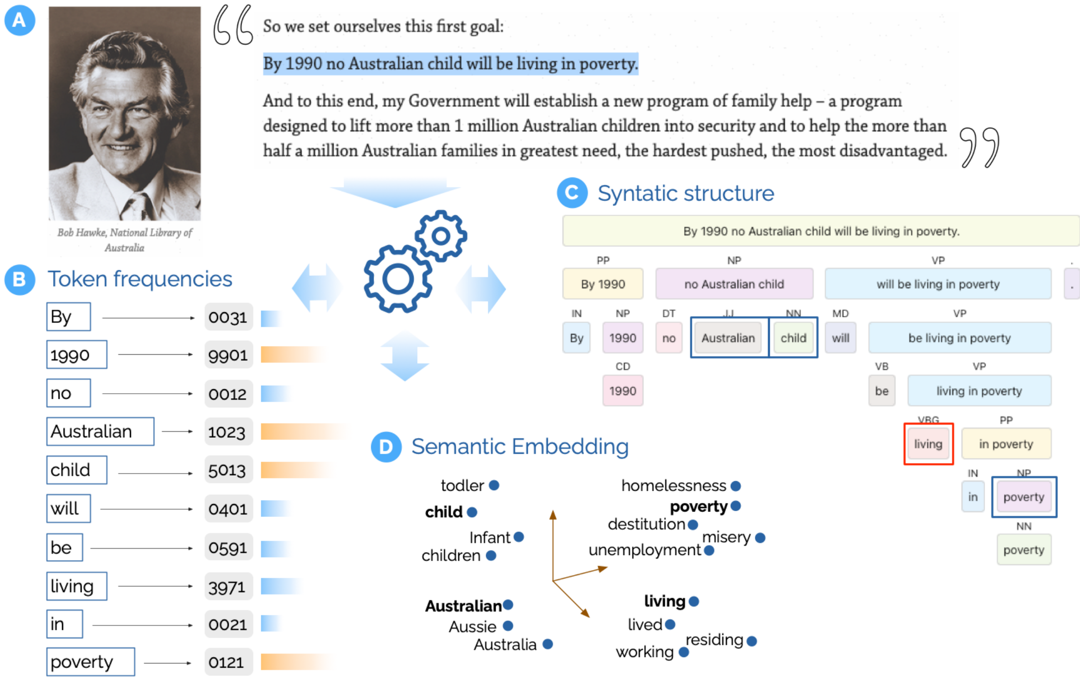

Figure 1. **Traditional (pre-transformer) NLP approaches to analysing discourse.** *Given some text (A), unique words (tokens) can be counted and turned into frequencies (B), syntactic structures can be identified and named-entities can be labelled (C), and words can be associated with their location in high-dimensional, pre-trained, semantic embedding spaces (D). .*

of discourse artefacts this has created (Mauro et al., 2016), has seen a move to automation of the standard approach, in particular, Step 5 – the coding of texts. Yet, as will be shown below, automating, and so scaling, the standard approach to framing analysis is just the starting point for automated methods with new technological frontiers opening up new ways of analysing and tracking public discourse.

Here, the fundamental challenge of computation lies at the heart of technological advance: how to encode the richness of human semantics in machine-readable format? Computational linguistics is the field of computer science perhaps most associated with attempting to answer this question, and multiple techniques have arisen over recent times which are still in use by computational social-, political-, and economic- scientists alike, to study public discourse (Bail, 2014; Gentzkow et al., 2019).

Indeed, we can chart the trajectory of NLP research which, starting with proximate abstractions of narrative framing analysis (e.g. word occurrences, topic modelling, semantic analysis), has been narrowing closer towards the object itself – trying to identify accurately elements of speech which encode one framing or another (rich, nuanced and contextual) in text, at scale. Along the way, several ingenious approaches to quantifying discourse dynamics aside from simply 'coding' has arisen, yielding insights derived from much larger volumes of text than could ever have been feasibly handled by human means alone.

In figure 1 we summarise the fundamental technologies used in traditional (pre-transformer) NLP research applications in the social science. We discuss each in turn.

Token frequencies

– A classic NLP approach is to define a unique vocabulary for a corpus of texts, typically dropping short, non-meaningful words - 'stop-words' ('a', 'to', 'is', 'the', etc.), and reducing words to their root or stem form (e.g. 'running' → 'run') so as to create a minimal set of meaningful word forms, known

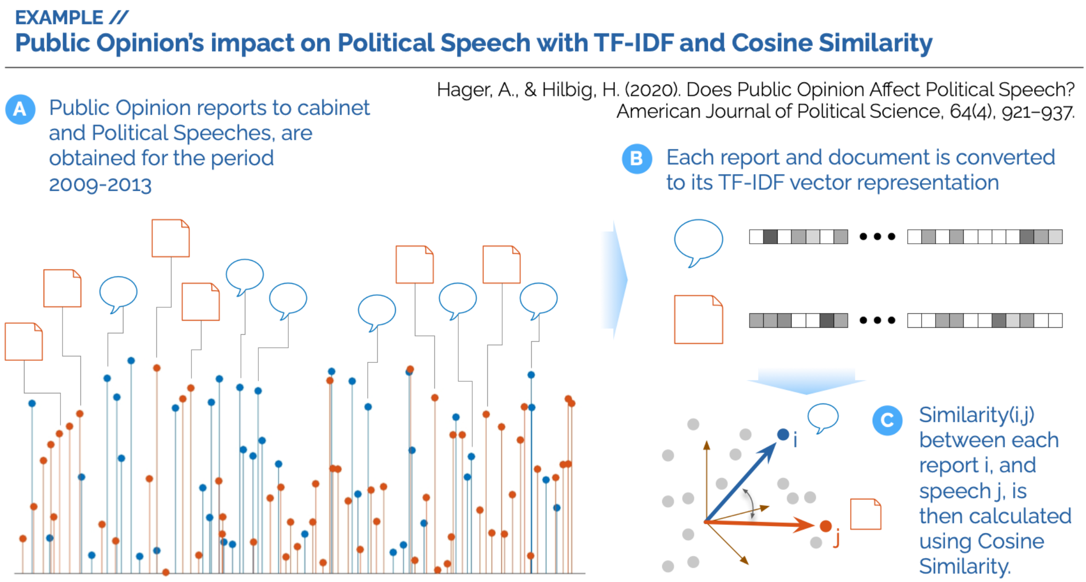

Figure 2. **Causal analysis of public opinion's impact on political discourse.** *In Hager and Hilbig (2020) a large corpus of German cabinet public opinion surveys were obtained after release, and, together with political speeches, press releases and other announcements of cabinet members (A), a regression discontinuity causal analysis framework was undertaken. After standard pre-processing, 3,860 length tf-idf vectors for each report and speech were calculated (B), enabling the calculation of semantic similarity between each report—speech pair via cosine similarity (C)..*

as 'tokens', by which any text can be represented. The computational task is then to count up the number of times each token appears in a given text ('token frequencies'). Importantly, such frequency representations lose any *contextual* information about the token. As such, these representations are called 'bag-of-word' approaches, since it is as if all of the words have been cut out of their context and tipped into a bag, regardless of order. Whilst such an approach may seem overly reductive, further processing can recover remarkable insights. For instance, by converting token frequencies ('tf') for a text into a vector, and then modifying token frequencies by a scarcity weighting for that token across all texts ('idf', inverse document frequency), researchers can then compute semantic distances between texts in this vector space.

An example of such an approach is found in Hager and Hilbig (2020) who analyse public opinion survey texts along with political speeches to examine the causal relationship between public opinion and political discourse by calculating tf-idf vector distances (see Fig. 2). Alternatively, tf-idf vectors can be fed into topic-modelling algorithms which perform unsupervised clustering[5] over the vectors typically using latent Dirichlet allocation (LDA) methods, such that topic dynamics can be followed over time (Jelodar et al., 2019). Examples of this strategy include Vidgen and Yasseri (2020b) who study the impact of petitioners on policy direction in the United Kingdom, Bonilla and Mo (2019) who analyse discourse topics in newspapers related to human trafficking, and Goyal and Howlett (2021) who study Covid-19 policy responses.

Other extensions to token frequency approaches include counting specific tokens which are said to encode attributes of language like sentiment. So-called 'dictionary-based' sentiment classifiers such as the Lexicoder Sentiment Dictionary (Young and Soroka, 2012) or VADER - Valence Aware Dictionary for Sentiment Reasoning (Hutto and Gilbert, 2014) are two examples of such approaches.

Whilst tf-idf methods are reasonably suited to topic modelling, the fact remains that since tf-idf vectors treat language use as entirely context-free, bag-of-word approaches are inherently limited in their ability to accurately identify nuanced semantic patterns which are inherent in narrative discourse analysis. Sentiment analysis being a good example of such limitations – dictionaries of terms associated with positive or negative sentiment are very labour-intensive to build, are domain specific, and are entirely context free, making sarcasm or irony a major stumbling block in their application (Grimmer and Stewart, 2013).

Semantic embeddings

– A major revolution occurred in NLP at the start of the second decade of the millennia, with the development of massively trained neural word-embedding models. Here, neural network architectures are trained to guess the correct token from surrounding tokens, or alternatively, the surrounding tokens from a given token. For the first time, computational representation of human language went beyond mere frequencies of unique index values, to actually associating words with their meaning, as revealed by their (local) context of use. Such models are referred to as 'word embedding' models (or sometimes, simply, 'embeddings'), with Word2Vec (Mikolov et al., 2013), GloVe (Pennington et al., 2014), and FastText (Bojanowski et al., 2017) among the most popular and widely used embeddings. By passing these models billions of sequences of human text (typically harvested from online sources), the models are able to accurately represent each word with a vector location in high-dimensional space. For instance, the vector for 'child' will be close - in vector space - to 'toddler', 'infant', and 'children'.

For the practitioner, word embeddings are very easy to accommodate, since any word (including stop-words; without stemming) can be passed to a pre-trained embedding (essentially a neural model), which deterministically produces a vector for that word. By repeating this process for a sequence of words in a text, many vectors are obtained, which can then be averaged to obtain a document location in 'semantic space'. From here, down-stream tasks like topic-modelling can be performed as introduced earlier.

However, there are more elaborate uses of word embeddings. A researcher can take the generically trained word-embedding model, and then further train the model on a sufficiently large sample of study texts, a process known as 'fine-tuning'. In this way, the word-embedding model itself starts to represent more closely the way that words encode semantics as revealed by the language in the study materials itself. An example of this approach is found in Kozlowski et al. (2019)'s ground-breaking study of class, where millions of book texts from each decade spanning the 1900s to the 1990s were each fed into a initialised Word2Vec model to fine-tune it to the particular semantic relationships of that decade. The authors then looked at semantic similarity between certain words which denote class and other words which denote dimensions of culture (see Fig. 3). Whilst not directly assessing framing within narrative discourses, one can see how deep cultural associations – the underpinning of narrative framing (ala Polletta and Callahan (2017)) – can be quantified and tracked over time using such an approach.

Syntactic structure

– Whilst grammar, on its own, does not closely associate with one narrative or frame, syntactic analysis empowers research methods to create more nuanced representations of narratives. Today, the field is almost entirely dominated by transformer methods, but there are still use-cases for traditional approaches, especially where data are scarce or computational resources are constrained. Here, hidden Markov Models or their derivatives represent pre-transformer approaches to labelling parts of speech McCallum et al. (2000); Brants (2000). Alternatively, other methods have been developed to correctly identify who is speaking in a sentence where pronouns replace names. This is especially important when conducting sentence-level analysis (a common approach in NLP), especially in political analysis, since knowing that the 'they' refers to the government, or opposition, or 'she' refers to a shadow-minister or the Prime Minister are critical pieces of information for any quantitative assessment. In NLP this task is referred to as 'coreference resolution' and fast, accurate methods are in common use (Clark and Manning, 2016a,0).

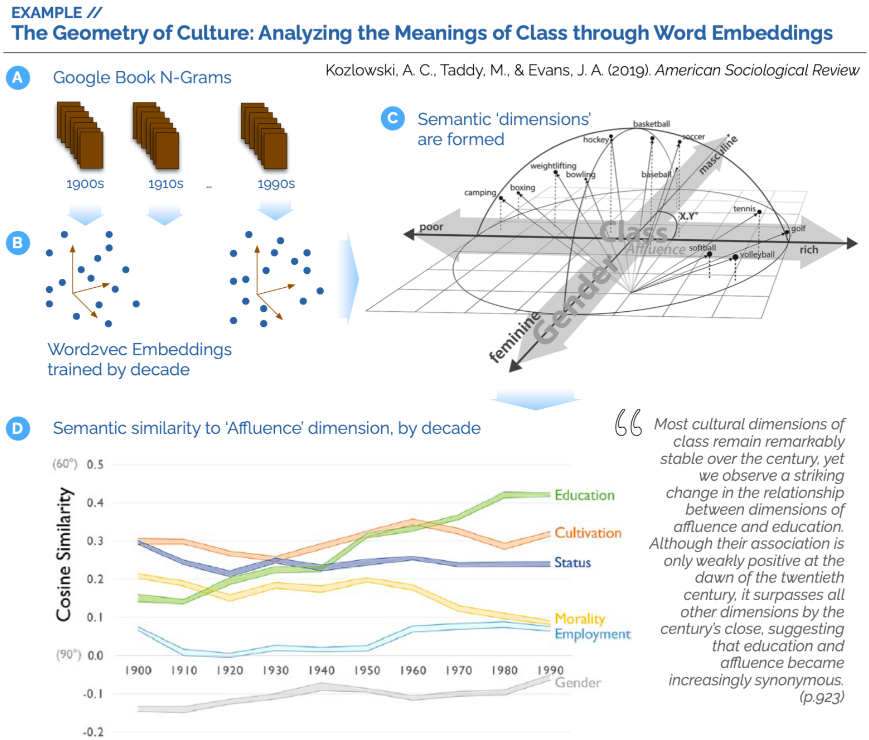

***Figure 3.* Using word-embeddings to analyse the foundations of narrative framing in millions of texts.** *Kozlowski et al. (2019) take US provenance Google Books n-gram texts, per decade (A), to fine-tune Word2Vec neural embedding models (B), such that distances between semantic dimensions along word-pairs lines in each decadal model can be calculated (C), to track the cultural association of e.g. 'Affluence' with these dimensions over time (D)..*

3.3. Ensemble approaches

In practice, traditional NLP workflows typically combine tools together in a sequence of steps, known as a 'pipeline' – a *combination* of transformations is used to achieve higher accuracy on a given research objective. This is the approach taken to a series of studies in computer science literature which define the task as a multi-class classification 'framing' problem[6]. However, this literature is of limited application to tracking conceptual framing (in the spirit of Entman, i.e. narratives to *persuade* or *promote*) since 'framing' is used in this literature to denote a *dimension* of an *issue* rather than what might be called the conceptual framing itself. For example, Boydstun et al. (2013)'s 15 generic policy 'frames' include 'Economic', 'Public opinion', and 'Cultural identity' – these are generic dimensions of an issue and do not carry enough meaning to denote a specific narrative or conceptual framing. In the literature these 'frames' are then instrumented across multi-issue labelled datasets such as the Media Frames Corpus (MFC) (Card et al., 2015) and the Gun Violence Frame Corpus (GVFC) (Liu et al., 2019) and analysed using a variety of multi-class classification algorithms (Field et al., 2018; Zhang et al., 2023).

However, not all computational literature follows this path, with several studies taking the more nuanced approach to framing analysis required for tracking narratives of democratic resilience. For example, Vidgen and Yasseri (2020a) offer a promising approach to tracking nuances in Islamic hate-speech in 4,000 social media posts. Namely, they distinguish between 'strong' and 'weak' islamophobia – a distinction which can be considered as two frames within the general attitude of islamophobia. In the study, they trial seven different pipelines, leveraging all of the traditional NLP approaches mentioned thus-far: tf-idf vectors, word embeddings, fine-tuned word-embeddings, sentiment, and syntactic features. They also trial a number of different classification algorithms from logistic regression, Naive-Bayes, through random forests, and support vector machines (SVM) and even deep learning (neural methods). The most accurate approach they identify combines features, and uses the SVM classifier, obtaining 72.17% accuracy on their training data ($n = 1,341$), and 77.3% accuracy on a further set of 300 hand-labelled test tweets. Likewise, Morstatter et al. (2018) study narratives of support and opposition (polarity) across 10 framings related to the balistic missile defence (BMD) issue in Europe. However, through hand-coding, and with training a multi-class NLP classifier, accuracy is low (less than 0.5 across classes).

Closer still to the narrative analysis task are studies which seek to track the incidence of 'micro-narrative' phrases within a large body of texts. Here, the work of Ash et al. (2024) and Anantharama et al. (2022) seek to identify 'Entity – Verb – Entity' (EVE) triplets, through a multi-step pipeline using word-embeddings (Ash et al.) or transformer-based sentences embeddings (Anantharama et al.), together with syntactic parsing and co-reference resolution. These approaches enable the tracking of phrases like 'God–bless–America' or 'indigenous voice–enshrine–constitution'. By clustering a set of these micro-narratives and associating with a particular framing, narrative components can be tracked across speakers, and through time.

So how well do non-transformer methods perform at traditionally hand-coded approaches to tracking discourse? A study by Nelson et al. (2021) considers three traditional methods (dictionary-based, tf-idf supervised, unsupervised topic discovery) applied to a large number of hand-coded news articles with a hierarchical identification schema relating to inequality. They find that tf-idf style classifiers were the most accurate approach in both identifying articles and tracking themes over time. However, with a combined precision/recall (f1) score of around 0.8, it would be hard to conclude that 'human-level accuracy' was obtained by the best method. However, the authors took methods 'off the shelf' and it is likely that more performance could be obtained by skilled practitioners.[7]

3.4. Harnessing the AI–transformer revolution

Sometime during 2016 or 2017 work began in the Google Research and Google Brain research groups on a revolutionary neural approach to the fundamental machine—human language problem. Known as the 'transformer' architecture, and introduced by the now famous paper entitled, 'Attention is all you need', Vaswani et al. (2017) demonstrated how, with vast quantities of human language training data, the new architecture could be taught, via the simplest of tasks – guess the next word in the sequence – to apparently infer complex linguistic and semantic properties of human speech. Subsequent work followed at Google and other labs, extending, refining, and enlarging the scale of the approach, resulting in breathtaking performance by transformer models on standard NLP benchmarks (Devlin et al., 2019; RaffelColin et al., 2020; Brown et al., 2020).

In Fig. 4 a generalised schematic of the transformer model architecture is shown. As neural models (i.e. large matrix multiplication models), transformer models take input 'at the left' and pass the transformed matrices of information from 'left to right' to arrive at a statistical prediction for the most likely word (token) that might come next. The model is trained by passing it billions of tokens from real sequences of human authored text obtained from online repositories of news, books, and social media, but hiding (or 'masking') the next word from the model. The model's guess for that next word (token) is then compared to the real (ground-truth) word (token) and the internal model weights are adjusted accordingly to nudge it towards more accurate next word prediction.

Importantly, these models can be auto-regressive in nature, that is, they can read their own produced tokens as input. This means that rather than a single next word prediction, they can be set to repeatedly generate next words, feeding the expanding sequence back into their input at each turn. This makes them truly 'generative' models, and has led to a large number of applications to traditionally complex NLP tasks. What once required a complex, multi-step pipeline to achieve, can now be achieved by

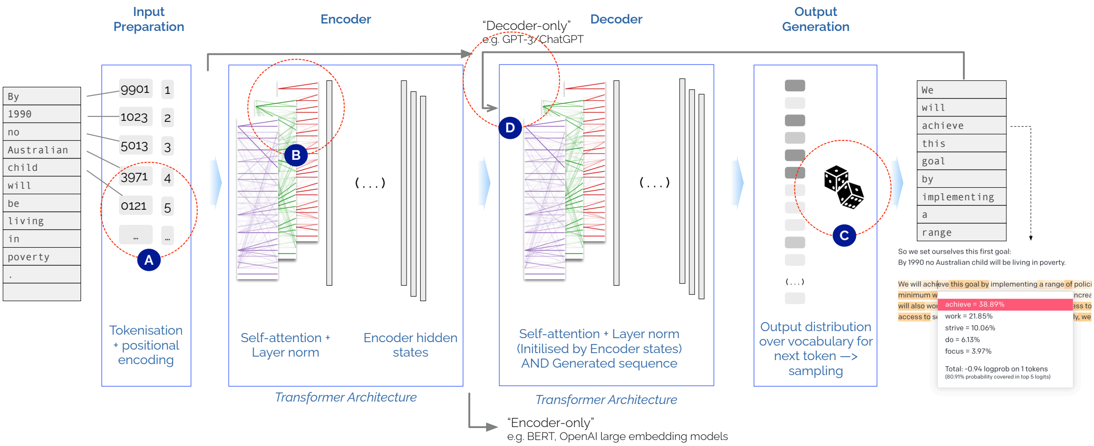

Figure 4. **Schematic of the large-language-model (LLM), encoder-decoder architecture.** *For the first time in NLP technology, text is fed into the model as a* sequence *(A), with positional encoding ensuring that subsequent steps retain knowledge of this ordering; transformer blocks are then applied in encoder, and/or decoder layers which enable the model to simultaneously attend to many aspects of the text (e.g. pronouns, verbs, nouns, punctuation) (B); the final layer is a statistical prediction for the next word (token) (C); and, each new token can be added to the input sequence, such that the model is able to keep generating new words (tokens) based on the generative steps already undertaken (D)..*

simply describing the task to the model in plain English, and subsequently passing the textual record (phrase, sentence, paragraph, or even whole document) to analyse to the model, and having the model 'generate' the information, label, or decision as desired.

Perhaps most remarkably, recent LLMs have been shown to exhibit a kind of *emergent* intelligence, performing tasks that cannot possibly have been part of their training data, even on complex logical problems hitherto thought of as well beyond the reach of 'thinking machines'. For example, Trinh et al. (2024) demonstrate 'AlphaGeometry' – a LLM trained on millions of examples of Euclidean plane geometry – which is able to perform close to gold medal (human) performance outcomes on the latest International Maths Olympiad problem set. Similarly, Bubeck et al. (2023) demonstrate OpenAI's GPT-4 (General Pre-trained Transformer, 4) ability to stack physical objects, draw unicorns (via coding), compose music and write complex code. Whilst still a topic of debate in the field of computer science, there are indications that recent LLMs have some kind of 'world model', that is, despite "simply" next token predicting, the model's training has somehow conferred some degree of knowledge of how physical objects and abstract concepts relate in real worlds, perhaps even encoding so-called 'common knowledge' (Mitchell and Krakauer, 2022; Mitchell, 2023).

Can these new technologies be turned to the challenges of tracking narratives in public discourse at scale? We consider three promising applications.

Transformers as labellers

– Perhaps the most obvious first application of the power of LLMs to understand and encode language, is to apply them to the labelling task ('step 5' in the Standard Approach, above), at scale. Here, it has become very apparent that LLMs are extremely good at this task, outperforming both expert and crowd sourced human labelling. Törnberg (2023)'s 2023 study shows that OpenAI's ChatGPT-4 achieved not only higher accuracy (>0.90 vs. 0.82), but also much higher inter-rater reliability (>0.95 vs. 0.65) than expert human coders, yet with statistically similar levels of bias.[8] Indeed, small amounts of human-labelled data can now be used to directly train LLMs towards higher accuracy in labelling large datasets very cheaply (Wang et al., 2021), or to 'teach' simple models to do the same, lifting their

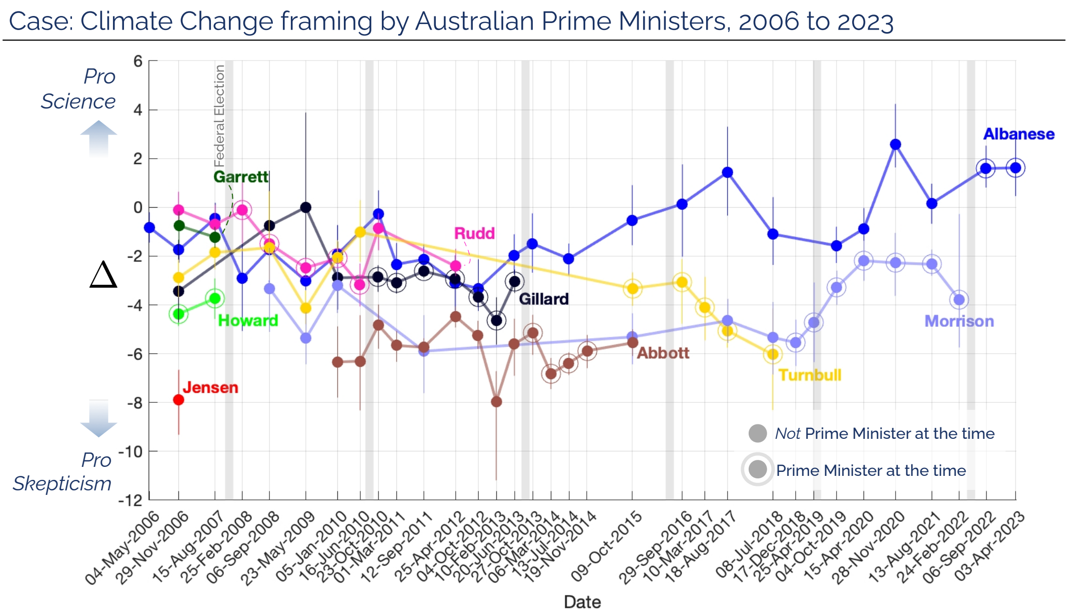

Figure 5. **Framing analysis with LLMs: 'paired-completion' (Angus and O'Neill, 2024) applied to climate change science framing in Australian Federal Parliament (2006-2023).** *Each line represents paired completion scores on a scale from negative (pro skepticism) to positive (pro science) for the speaker shown. Jensen and Garrett are added to the pool of Prime Ministers as they are well-known climate science denialist / supporters respectively. Rings around markers indicate that the speaker was Prime Minister at the time..*

performance in a process likened to slingshotting a space probe around a larger planet to go further, faster (O'Neill et al., 2023).

Transformers as synthetic data generators

– Alternatively, instead of dramatically reducing the cost of labelling real data, LLMs can be used to create large, synthetic textual datasets, with a highly controllable form. The benefit here is that LLM systems can be evaluated for accuracy, bias, and reliability on large amounts of labelled synthetic data where either real data are unavailable, or are very costly to curate, meaning that a system can be deployed on real data after significant refinement has already occurred on synthetic data. Techniques here focus on enforcing structured ouputs (Willard and Louf, 2023), or reducing hallucinations (Ayala and Béchard, 2024). However, recent advances by LLM platforms have meant that model control is now even more precise, with the ability for models to provide exact structured outputs fitting a researcher's schema (Pokrass, 2024).

Transformers as direct framing analysers

– Finally, recent research has shown that LLMs can be used directly as highly capable 'language models'. That is, that LLMs can be used to identify issue-framing in text directly, with low cost, and low bias. Angus and O'Neill (2024) introduce 'paired completion', a technique whereby LLMs are primed with examples of statements from one framing or another on a particular issue, and then the text to be analysed is fed as a 'completion' to the model. The method collects the likelihood that the model would have generated the completion text, conditional on already having 'said' the priming text. By comparing the completion likelihoods, a quantitative metric is obtained for the text's *textual alignment* with one frame or another. The study reports balanced accuracy scores as high as 0.93 across three synthetic narrative datasets covering climate change, domestic violence and misogyny. Figure 5 shows an example application of the method on the issue of climate change to a subset of speakers (mainly Prime Ministers) in Australian Federal Parliament, over 2006 to 2023 ($n = 4,606$ speeches). Despite requiring only a handful (around five) examples of pro-science and pro-skepticism texts to mark out each framing of the issue, paired-completion is able to accurately track the framing alignment for each speaker over two decades.

Alternatively, LLMs can encode an entire 'world' of a given source such as a particular newspaper or online media outlet (Guo et al., 2022). The approach here is to *fine-tune* a LLM, a method which trains a LLM to learn the particular patterns of speech of a given source corpus, by feeding it (say) news from one outlet only. In this way, a series of fine-tuned LLMs can be trained, one per news source, and then each LLM can be analysed for how it might complete a sentence, or use a key word in a sentence (e.g. complete the sentence, 'Immigrants travelling by boat to Australia are a ___ to our society.')

Using LLMs for framing analysis at scale is not without concerns. Generative LLMs present outstanding opportunities for quantified approaches to tracking narratives at scale, but their effective use requires a robust understanding of a suite of risks (Bommasani et al., 2021). Attention must be paid to the provenance of training data, bias in generated text, processing cost, and open or closed access, given that many of the most performant models are closed-source, run by large platforms. As mentioned, the training of LLMs (next word prediction) does not guarantee correctness or completeness in responses obtained, and even the most frontier LLMs can still be tripped up by seemingly trivial completion tasks[9].

4. Conclusion & Recommendations: the Future of Tracking Public Narratives of Democracy

Our starting point was that story-telling matters deeply to the way that we communicate and encode our culture and beliefs in society, and so, for policymakers concerned with democratic resilience, narrative analysis at scale is now an immensely valuable capability alongside traditional methodologies. We have seen that in social science, narratives can be analysed and tracked in a variety of ways, with issue-framing being one of the most common approaches. However, coding texts for the presence or absence of one or other framing has traditionally been a laborious process for experts and research assistants alike. And, human approaches are not without their own bias and reliability concerns. For this reason, we have considered how automated techniques arising from the computational sciences have been developed to analyse discourse in a variety of ways, including automation of the labelling task. Significantly we have seen that the most recent wave of technological advance, the wave that is sweeping every major discipline of the academy, the generative AI/LLM wave, holds enormous promise for application to tracking narratives with relatively low costs, low bias, and high reliability.

So how could these frontier techniques be applied to policy research questions relating to democratic resilience? We propose a number of starting points below.

4.1. Recommendations for data: removing blockers to policy-relevant discourse analysis at scale

Policymakers who wish to analyse the strength or fragility of our democracy by tracking narratives of democratic resilience, may rightly consider starting with the narratives spun by our own elected ministers, and the election platforms would-be ministers are touting. Together, these would provide a key input to narrative analysis for policy-relevant questions such as: is populist language on the rise? what fraction of election platforms in the recent election touted exclusionary nationalism? which jurisdictions are on the receiving end of political platforms that portray an 'Us vs. Them' narrative?

Unfortunately, through an analysis of the situation in several apparently healthy democracies, e.g. the United Kingdom (scoring 91 on Freedom House's 'Global Freedom Scores'[10]), Australia (95), and Finland (100), we find that, despite the promise of NLP/AI technology, answering these questions is still frustrated by failures in rather more basic requirements.

4.1.1. Data Recommendation 1: an official API for Parliamentary Hansard

Contrary to what are likely reasonable public expectations, the public discourse that one would *most* expect to be freely available under well documented and supported means, namely Parliamentary Hansard, is not, in fact, easy to come by for the kind of analysis proposed in the current paper. The lack of a single, comprehensive access point for Parliamentary discourse data appears to bedevil efforts across jurisdictions. Key to such access is support for an application programming interface (API). As opposed to browsing, where an individual clicks on links and reads the content of a webpage visited, an API enables scripted (i.e. computational) access to a database, which any down-stream, large-scale, and real-time analysis of public discourse requires.

Efforts vary by country but the pattern is the same. In the United Kingdom, after a historical period of access[11], the 'data.parliament.uk' initiative hosts a series of APIs to access various aspects of UK Parliamentary activity [12]. However, the API is still in 'beta' mode, and does not support access to Parliamentary Debate Hansard. The situation in Australia is worse. Whilst each jurisdiction has some user-facing web-portal for Hansard access[13] these require laborious human operation to download a large number of speeches. To our knowledge only the NSW Parliament has created an API access point to their Hansard, hosted via data.gov.au[14] which demonstrates that the task is surmountable. Likewise, Finland provides no official API to its Hansard, publishing instead, links to sessions of Parliamentary plenaries, including webcasts[15].

Into the breach, community funded, not-for-profit, projects have arisen to make parliamentary Hansard available. Typically, these efforts effectively reverse-engineer the task by visiting the user-facing (non-API) Hansard pages, downloading the page, and then transforming the information from the page back into a database to be finally made available on a community funded API portal. Naturally, these efforts, although well-intentioned and often achieving a high level of reliability, are not assured, and are not an 'official' source of political speech. For instance, the 'TheyWorkForYou'[16] initiative in the UK provides such a service, whilst in Australia, the 'Open Australia' initiative[17] tries to provide this access, but is typically not up to date, struggling to get the man-power to provide reliable service. Likewise in Finland, community scraped APIs are available with methods which unlock them, but these projects do not appear to be up-to-date[18].

Clearly, for any major policy-relevant democratic resilience analysis of Parliamentary discourse, or for the building of tools on top of Hansard to drive transparency, accountability, and public understanding, enabling civil society and academia to process the very speeches of our democracy is a critical starting point.

4.1.2. Data Recommendation 2: an official repository of political platforms

Relatedly, there are rarely official comprehensive sources for election literature, speeches, or materials. In Australia, whilst the National Library offers a portal[19], following links is again a laborious task and not designed for programmatic access. Digital election materials are available through the 'Trove' web archive site, however, links back to state Hansard are often dead due to stale or broken connections. In the UK, the Electoral Commission is tentatively exploring API development, with an 'alpha' version providing some information on candidates (name, party affiliation), though this API is itself delivered by a not-for-profit, 'Democracy Club', which hosts a more thorough API which can provide, among other things, candidate information including 'statements to voters' ("where collected")[20].

To be sure, there are promising international initiatives in this direction, however, their coverage is typically restricted to major political platforms, at the National level, befitting an international research initiative. For example, The Political Party Database Project[21] provides a smattering of official party documents for Australia, the United Kingdom, and Finland's major parties (e.g. in Australia: Labor, Liberal, National, Greens) over the last two decades. Similarly, the highly regarded Manifesto-Project[22] provides a small sub-set of national party platforms for each country, albeit with a much greater longitudinal coverage. Yet, the choice of which party platform to include, and even which document is stored, is down to academic discretion and resourcing. Without official support, these databases miss the small, tail-end party platforms, which arguably have the most potent information for policymakers wishing to track emergent narratives of democracy.

To facilitate open and active engagement by the research community with the platform and positional materials of the full range of democratic actors, programmatic (API) end-points of political platforms should be created in the ideal case, or at least a location enabling download with filtering should be made, so that large-scale textual research can be realised.

4.2. Recommendations for future research: towards real-time tracking of policy-relevant democratic discourse

We finish by sketching a potential agenda for further democratic narrative research that would support policymakers, academic researchers, and the wider community alike. Granted, significant data challenges exist, but laying these to one side, we consider a world where data were more available, enabling enriched analysis with AI-Transformer backed technologies considered earlier.

4.2.1. Research Recommendation 1: Hearing from the fringe – augmenting surveys with alternative data

LLM-backed methods offer a chance to overcome two key challenges with traditional, survey-based methods of tracking anti-democratic narratives at nation-scale. First, surveys are typically expensive to run, and therefore are undertaken at most annually, though often less frequently. Second, a common concern of those who run surveys of trust in government and institutions, is that, despite rigorous attempts, survey respondents are typically self-selecting. That is, if there is a growing group of people in a country who support anti-democratic narratives, they are also highly likely to be wary of institutional mechanisms in general, and so, unlikely to be willing to participate in surveys run by 'official' organisations. The UK Government, for example, acknowledge this source of bias explicitly in their 'Trust in Government Survey QMI' report (2024)[23]

> There may be a sampling bias, because the sampling frame comes from those who have already responded to a government survey; these people may differ to the general population in how much they trust the government and other public institutions.

However, alternative data sources abound in these countries, from online discussion-boards, to YouTube and Instagram content, to broad-based social media and alt- group websites. By converting these sources to text (admittedly, not an insignificant task in some cases), LLM-enabled framing analysis on dimensions of populism, 'Us vs. Them' and other anti-democratic narratives can be tracked. By building up a time-series of these markers, movements up and down in the concentration and strength of these narratives can be compared statistically to baseline levels. Significant movements could trigger more intensive survey or focus-group work, or deeper dive analysis of specific narratives and the basis of their claims. In this way, the 'fringe' of society who may be actively engaging in digital media and platforms, but who are already disengaged from government surveys and institutions, can be better understood in policy-relevant analysis of the state or health of democratic values in society at large.

4.2.2. Research Recommendation 2: Estimating rates of anti-democratic support with revealed preferences, the case of Australia

Australia's democracy is one of only a handful of free countries world-wide to enforce compulsory voting. With the measure's introduction in 1924, Australia will see 100 years of the practice through the election cycle of 2025. However, for similar selection issues discussed above, a significant challenge in Australia is estimating the *true* fraction of the population who hold more extreme anti-democratic views. Given that voter turn-out is very high across the country, and that voting is considered a private act, one can consider voter support for radical party positions as a more revealed preference indicator of support, than perhaps stated preference measures as considered above. Here, the challenge is to collect and analyse the thousands of political tracts from major and minor parties across all election cycles, at state and national level, and analyse these for narratives of democratic resilience and opposition. By applying vote-share to those party platforms that score strongly on textual alignment with particular narratives of democratic opposition, one can calculate, for the first time, a true level of revealed support for these positions. These data would go significantly towards answering the question of whether Australia really does have a 'hidden' underbelly of support for radical positions on democratic issues, or if instead, public posturing does not speak to privately held beliefs.

Whilst this research direction focusses on the case of Australia, there are over a dozen countries which share the compulsory voting approach, providing further contexts for the application of AI-transformer methodologies to track relevant narratives of democracy by jurisdiction.

4.2.3. Research Recommendation 3: Supporting accountability in our Parliamentary discourses: a public monitor of democratic discourse

One of the major challenges facing the public, journalists, and academics alike, is the challenge of identifying objectively that a speaker or outlet is saying something unusual (for them) or fitting a particular narrative of democracy. Without accurate, quantitative tools, we are left relying on 'opinion' pieces and expert talking heads to opine on whether a particular speaker's comments really were anti-democratic, populist, or divisive. Whilst expert commentary will always be important, what we lack is a way for the public to see insightful visualisations, driven by accurate discourse measurement, of how a particular discourse fits in to a spectrum of discourse at a point in time, or across time. Such a 'Narrative Observatory' like tool would be a major benefit to the democratic project. Speakers would be held immediately accountable for their words, policymakers, journalists and analysts would more easily be able to draw comparisons and make data-driven insights about the positions being taken on the floors of our Parliaments nationally. Fact-checking initiatives exemplify this kind of public good provision, a narrative observatory would complement and empower our society to better understand and hold to account those who represent us.

Notes

[1]In political and social sciences it is traditional to use 'coding' to refer to the identification of some attribute in a text, typically by a human reader. In machine learning and computer science more generally, this process is analogous to 'labelling' data. As such, we use these terms interchangeably.

[2]For a review of survey based methods, see surveys such as Levi and Stoker (2000); Citrin and Stoker (2018), or projects such as the Australian Social Cohesian index (https://scanloninstitute.org.au/research/australian-cohesion-index), or the Social Cohesion Radar (Dragolov et al., 2016).

[3]For example, very short utterances such as interjections in Parliament may be ambiguous, whilst very long speeches may cover many issues and so are too diffuse in their connection to the study.

[4]See Krippendorff (2019) for an overview of evaluation and reliability methods, including e.g. 'Krippendorf's Alpha', commonly used for categorical agreement between coders.

[5]In machine learning, an algorithm is either trained with, or without, labelled data. In the former, known as 'supervised' learning, the objective is to accurately infer the class (label) of a given record, given that the actual class is known ('ground-truth'). In the latter, known as 'unsupervised' learning, the objective is to identify patterns in the data, minimising some penalty which increases if similar records are not grouped together.

[6]See Ali and Hassan (2022) for a review of 37 computational approaches in this direction.

[7]For an earlier survey of text as data methods see Grimmer and Stewart (2013), who reach largely the same conclusion.

[8]One can also leverage the *encoding* capability of LLMs to get better 'features' for multi-class labelling. This is the approach of Mendelsohn et al. (2021), working again with the 15 generic 'frames' of Boydstun et al. (2013). However, reported balanced accuracies are far lower (~ 0.55).

[9]For example, if one asks the most performant language model to date, Anthropic's Claude 3.5 Sonnet, 'How many i's in Kit-Kat?', it answers, 'The correct answer is that there are zero i's in Kit Kat.' This behaviour is common for all the major models alike.

[10]Freedom House, https://freedomhouse.org/countries/freedom-world/scores.

[11]See the historical Hansard Millbank home for such data at https://api.parliament.uk/historic-hansard/api.

[12]See https://explore.data.parliament.uk/index.html.

[13]e.g. Federal Parliament provides a search front-end: https://www.aph.gov.au/Parliamentary_Business/Hansard/ and Victorian Parliament the same: https://www.parliament.vic.gov.au/parliamentary-activity/hansard/hansard-debate/, neither offer programmatic (API), or bulk down-load access directly.

[14]See: https://parliament-api-docs.readthedocs.io/en/latest/new-south-wales/.

[15]See https://verkkolahetys.eduskunta.fi/fi/ (Finnish/Swedish).

[16]See: https://www.theyworkforyou.com/api/.

[17]See: https://www.openaustralia.org.au/api/.

[18]See for example, https://github.com/rOpenGov/finpar.

[19]See: https://www.nla.gov.au/collections/what-we-collect/ephemera/federal-election-campaigns.

[20]See https://democracyclub.org.uk/data_apis/voting_information_api/ and links therein.

[21]See https://www.politicalpartydb.org/countries/.

[22]See https://manifesto-project.wzb.eu/.

[23]Office for National Statistics (ONS), released 1 March 2024, ONS website, methodology, Trust in Government Survey QMI. See {urlhttps://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/methodologies/trustingovernmentsurveyqmi

Acknowledgments. I am grateful to fellow members of the Resilient Democracy Research and Data Network (Australia) for discussions which informed the development of this paper. In particular, from the network, Prof Nicholas Biddle (Australian National University) who provided useful comments on earlier drafts of this work. All errors and omissions are my own.

Competing Interests. 'None'.

Data Availability Statement. NA.

Ethical Standards. The research meets all ethical guidelines, including adherence to the legal requirements of the study country.

Author Contributions. S.D.A. is the sole author of this manuscript and is responsible for the conceptualization, methodology, formal analysis, investigation, data curation, writing (original draft preparation and review/editing), visualization, and project administration.

Supplementary Material. NA.

References

Algan, Y., Guriev, S., Papaioannou, E., and Passari, E. (2017). The european trust crisis and the rise of populism. *Brookings Papers on Economic Activity*, pages 309–382.

Ali, M. and Hassan, N. (2022). A survey of computational framing analysis approaches. *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, EMNLP 2022*, pages 9335–9348.

Anantharama, N., Angus, S. D., and O'Neill, L. (2022). Canarex: Contextually aware narrative extraction for semantically rich text-as-data applications. *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 3551–3564.

Angus, S. D. and O'Neill, L. (2024). Paired completion: quantifying issue-framing at scale with LLMs. SoDa Laboratories Working Paper Series 2024-02, Monash University, SoDa Laboratories.

Ash, E., Gauthier, G., and Widmer, P. (2024). Relatio: Text semantics capture political and economic narratives. *Political Analysis*, 32:115–132.

Ayala, O. and Béchard, P. (2024). Reducing hallucination in structured outputs via retrieval-augmented generation. *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 6: Industry Track)*, 6:228–238.

Bail, C. A. (2014). The cultural environment: Measuring culture with big data. *Theory and Society*, 43:465–482.

Benier, K., Wickes, R., and Moran, C. (2020). 'african gangs' in australia: Perceptions of race and crime in urban neighbourhoods. *Journal of Criminology*, 54:220–238.

Bojanowski, P., Grave, E., Joulin, A., and Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135–146.

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., Donahue, C., Doumbouya, M., Durmus, E., Ermon, S., Etchemendy, J., Ethayarajh, K., Fei-Fei, L., Finn, C., Gale, T., Gillespie, L., Goel, K., Goodman, N., Grossman, S., Guha, N., Hashimoto, T., Henderson, P., Hewitt, J., Ho, D. E., Hong, J., Hsu, K., Huang, J., Icard, T., Jain, S., Jurafsky, D., Kalluri, P., Karamcheti, S., Keeling, G., Khani, F., Khattab, O., Kohd, P. W., Krass, M., Krishna, R., Kuditipudi, R., Kumar, A., Ladhak, F., Lee, M., Lee, T., Leskovec, J., Levent, I., Li, X. L., Li, X., Ma, T., Malik, A., Manning, C. D., Mirchandani, S., Mitchell, E., Munyikwa, Z., Nair, S., Narayan, A., Narayanan, D., Newman, B., Nie, A., Niebles, J. C., Nilforoshan, H., Nyarko, J., Ogut, G., Orr, L., Papadimitriou, I., Park, J. S., Piech, C., Portelance, E., Potts, C., Raghunathan, A., Reich, R., Ren, H., Rong, F., Roohani, Y., Ruiz, C., Ryan, J., Ré, C., Sadigh, D., Sagawa, S., Santhanam, K., Shih, A., Srinivasan, K., Tamkin, A., Taori, R., Thomas, A. W., Tramèr, F., Wang, R. E., and Wang, W. (2021). On the opportunities and risks of foundation models.

Bonilla, T. and Mo, C. H. (2019). The evolution of human trafficking messaging in the united states and its effect on public opinion. *Journal of Public Policy*, 39:201–234.

Boydstun, A. E., Gross, J. H., Resnik, P., and Smith, N. A. (2013). Identifying media frames and frame dynamics within and across policy issues. In *Proceedings of New Directions in Analyzing Text as Data Workshop*. Part of the New Directions in Analyzing Text as Data Workshop.

Brants, T. (2000). TnT – a statistical part-of-speech tagger. In *Sixth Applied Natural Language Processing Conference*, pages 224–231, Seattle, Washington, USA. Association for Computational Linguistics.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., and Amodei, D. (2020). Language Models are Few-Shot Learners. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M. F., and Lin, H., editors, *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc.

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M. T., and Zhang, Y. (2023). Sparks of artificial general intelligence: Early experiments with gpt-4.

Card, D., Boydstun, A. E., Gross, J. H., Resnik, P., and Smith, N. A. (2015). The media frames corpus: Annotations of frames across issues. In Zong, C. and Strube, M., editors, *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 438–444, Beijing, China. Association for Computational Linguistics.

Chong, D. and Druckman, J. N. (2007). Framing theory. *Annual Review of Political Science*, 10:103–126.

Citrin, J. and Stoker, L. (2018). Political trust in a cynical age. *Annual Review of Political Science*, 21:49–70.

Clark, K. and Manning, C. D. (2016a). Deep reinforcement learning for mention-ranking coreference models. *EMNLP 2016 - Conference on Empirical Methods in Natural Language Processing, Proceedings*, pages 2256–2262.

Clark, K. and Manning, C. D. (2016b). Improving coreference resolution by learning entity-level distributed representations. *54th Annual Meeting of the Association for Computational Linguistics, ACL 2016 - Long Papers*, 2:643–653.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Dragolov, G., Ignácz, Z. S., Lorenz, J., Delhey, J., Boehnke, K., and Unzicker, K. (2016). *Social Cohesion in the Western World: What Holds Societies Together: Insights from the Social Cohesion Radar*. Springer.

Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, 43:51–58.

Field, A., Kliger, D., Wintner, S., Pan, J., Jurafsky, D., and Tsvetkov, Y. (2018). Framing and agenda-setting in russian news: a computational analysis of intricate political strategies. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP 2018*, pages 3570–3580.

Gentzkow, M., Kelly, B., and Taddy, M. (2019). Text as data. *Journal of Economic Literature*, 57:535–74.

Gidron, N. and Hall, P. A. (2020). Populism as a problem of social integration. *Comparative Political Studies*, 53:1027–1059.

Goyal, N. and Howlett, M. (2021). "Measuring the Mix" of Policy Responses to COVID-19: Comparative Policy Analysis Using Topic Modelling. *Journal of Comparative Policy Analysis: Research and Practice*, 23(2):250–261. Publisher: Routledge.

Grimmer, J. and Stewart, B. M. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, 21(3):267–297. Edition: 2017/01/04 Publisher: Cambridge University Press.

Guo, X., Ma, W., and Vosoughi, S. (2022). Capturing topic framing via masked language modeling. *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 6811–6825.

Hagen, J. and Lysne, O. (2016). Protecting the digitized society-the challenge of balancing surveillance and privacy. *The Cyber Defense Review*, 1:75–90.

Hager, A. and Hilbig, H. (2020). Does public opinion affect political speech? *American Journal of Political Science*, 64:921–937.

Hetherington, M. J. and Rudolph, T. J. (2015). *Why Washington won't work : polarization, political trust, and the governing crisis*. The University of Chicago Press.

Holloway, J. and Manwaring, R. (2023). How well does 'resilience' apply to democracy? a systematic review. *Contemporary Politics*, 29:68–92.

Hutto, C. and Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 8. Issue: 1.

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., and Westwood, S. J. (2019). The origins and consequences of affective polarization in the united states. *Annual Review of Political Science*, 22:129–146.

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., and Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimedia Tools and Applications*, 78(11):15169–15211.

Jennings, W. and John, P. (2009). The dynamics of political attention: Public opinion and the queen's speech in the united kingdom. *American Journal of Political Science*, 53:838–854.

Kozlowski, A. C., Taddy, M., and Evans, J. A. (2019). The geometry of culture: Analyzing the meanings of class through word embeddings. *American Sociological Review*, 84:905–949.

Krippendorff, K. (2019). Content analysis: An introduction to its methodology. *Content Analysis: An Introduction to Its Methodology*.

Levi, M. and Stoker, L. (2000). Political trust and trustworthiness. *Annual Review of Political Science*, 3:475–508.

Liu, S., Guo, L., Mays, K., Betke, M., and Wijaya, D. T. (2019). Detecting frames in news headlines and its application to analyzing news framing trends surrounding u.s. gun violence. *CoNLL 2019 - 23rd Conference on Computational Natural Language Learning, Proceedings of the Conference*, pages 504–514.

Martino, L. D. (2020). The spectrum of listening. *Routledge Handbook of Public Diplomacy*, pages 21–29.

Mauro, A. D., Greco, M., and Grimaldi, M. (2016). A formal definition of big data based on its essential features. *Library Review*, 65:122–135.

McCallum, A., Freitag, D., and Pereira, F. C. N. (2000). Maximum entropy markov models for information extraction and segmentation. In *Proceedings of the Seventeenth International Conference on Machine Learning*, ICML '00, page 591–598, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

McCoy, J., Rahman, T., and Somer, M. (2018). Polarization and the global crisis of democracy: Common patterns, dynamics, and pernicious consequences for democratic polities. *American Behavioral Scientist*, 62:16–42.

McCoy, J. and Somer, M. (2019). Toward a theory of pernicious polarization and how it harms democracies: Comparative evidence and possible remedies. *Annals of the American Academy of Political and Social Science*, 681:234–271.

Mendelsohn, J., Budak, C., and Jurgens, D. (2021). Modeling framing in immigration discourse on social media. *NAACL-HLT 2021 - 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference*, pages 2219–2263.

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. *1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings*.

Mitchell, M. (2023). Ai's challenge of understanding the world. *Science*, 382.

Mitchell, M. and Krakauer, D. C. (2022). The debate over understanding in ai's large language models. *Proceedings of the National Academy of Sciences of the United States of America*, 120.

Moore, A. D. (2011). Privacy, security, and government surveillance: Wikileaks and the new accountability. *Public Affairs Quarterly*, 25:141–156.

Morstatter, F., Wu, L., Yavanoglu, U., Corman, S. R., and Liu, H. (2018). Identifying framing bias in online news. *ACM Transactions on Social Computing*, 1:1–18.

Mylonas, H. and Tudor, M. (2021). Nationalism: What we know and what we still need to know. *Annual Review of Political Science*, 24:109–132.

Nelson, L. K., Burk, D., Knudsen, M., and McCall, L. (2021). The Future of Coding: A Comparison of Hand-Coding and Three Types of Computer-Assisted Text Analysis Methods. *Sociological Methods & Research*, 50(1):202–237.

Norris, P. and Inglehart, R. (2019). Cultural backlash: Trump, brexit, and authoritarian populism. *Cultural Backlash: Trump, Brexit, and Authoritarian Populism*, pages 1–540.

O'Neill, L., Anantharama, N., Borgohain, S., and Angus, S. D. (2023). Models teaching models: Improving model accuracy with slingshot learning. In Vlachos, A. and Augenstein, I., editors, *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 3233–3247, Dubrovnik, Croatia. Association for Computational Linguistics.

Penney, J. W. (2016). Chilling effects: Online surveillance and wikipedia use. *Technology Law Journal*, 31:117–182.

Pennington, J., Socher, R., and Manning, C. (2014). GloVe: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Doha, Qatar. Association for Computational Linguistics.

Pokrass, M. (2024). Introducing structured outputs in the api | openai.

Polletta, F. and Callahan, J. (2017). Deep stories, nostalgia narratives, and fake news: Storytelling in the trump era. *American Journal of Cultural Sociology*, 5:392–408.

RaffelColin, ShazeerNoam, RobertsAdam, LeeKatherine, NarangSharan, MatenaMichael, ZhouYanqi, LiWei, and J., L. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *The Journal of Machine Learning Research*, 21:1–67.

Richards, N. M. (2013). The dangers of surveillance. *Harvard Law Review*, 126:1934–1965.

Shiller, R. J. (2019). *Narrative Economics: How Stories Go Viral and Drive Major Economic Events*. Princeton University Press, Princeton, NJ.

Smith-Carrier, T. and Lawlor, A. (2016). Realising our (neoliberal) potential? a critical discourse analysis of the poverty reduction strategy in ontario, canada. *Critical Social Policy*, 37:105–127.

Snow, D. A. and Benford, R. D. (1988). Ideology, frame resonance, and participant mobilization. *International Social Movement Research*, 1:197–217.

Somers, M. R. (1994). The narrative constitution of identity: A relational and network approach. *Theory and Society*, 23:605–649.

Terkildsen, N. and Schnell, F. (1997). How media frames move public opinion: An analysis of the women's movement. *Politcal Research Quarterly*, 50:879–900.

Trinh, T. H., Wu, Y., Le, Q. V., He, H., and Luong, T. (2024). Solving olympiad geometry without human demonstrations. *Nature 2024 625:7995*, 625:476–482.

Törnberg, P. (2023). Chatgpt-4 outperforms experts and crowd workers in annotating political twitter messages with zero-shot learning.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). Attention is All you Need. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Vidgen, B. and Yasseri, T. (2020a). Detecting weak and strong islamophobic hate speech on social media. *Journal of Information Technology & Politics*, 17:66–78.

Vidgen, B. and Yasseri, T. (2020b). What, when and where of petitions submitted to the UK government during a time of chaos. *Policy Sciences*, 53(3):535–557.

Wang, S., Liu, Y., Xu, Y., Zhu, C., and Zeng, M. (2021). Want to reduce labeling cost? gpt-3 can help. *Findings of the Association for Computational Linguistics, Findings of ACL: EMNLP 2021*, pages 4195–4205.

Willard, B. T. and Louf, R. (2023). Efficient guided generation for large language models.

Young, L. and Soroka, S. (2012). Affective News: The Automated Coding of Sentiment in Political Texts. *Political Communication*, 29(2):205–231. Publisher: Routledge.

Zhang, Y., Shah, D., Pevehouse, J., and Valenzuela, S. (2023). Reactive and asymmetric communication flows: Social media discourse and partisan news framing in the wake of mass shootings. *International Journal of Press/Politics*, 28:837–861.