Virality:
What Makes Narratives Go Viral, and Does it Matter?*

Kai Gehring† and Matteo Grigoletto‡

September 19, 2025

Abstract

Understanding behavioral aspects of collective decision-making remains a core challenge in economics. *Political narratives* can be seen as a key communication technology that shapes and affects human decisions beyond pure information transmission. The effectiveness of narratives can be driven as much by their virality as by their specific persuasion power. To analyze *political narratives* empirically, we introduce the political narrative framework and a pipeline for its measurement using large language models (LLMs). The core idea is that the essence of a narrative can be captured by its characters, which take on one of three archetypal roles: hero, villain, and victim. To study what makes narratives go viral we focus on the topic of climate change policy and analyze data from the social media platform Twitter over the 2010–2021 period, using retweets as a natural measure of virality. We find that *political narratives* are consistently more viral than neutral messages, irrespective of time or author characteristics and other text features. Different role depictions differ in terms of emotional language, but *political narratives* capture more than merely valence or emotions. Hero roles and human characters increase virality, but the biggest virality boost stems from using villain roles and from combining other roles with villain characters. We then examine the persuasiveness of *political narratives* using a set of online experiments. The results show that narrative exposure influences beliefs and revealed preferences about a character, but a single exposure is not sufficient to move support for specific

*We thank Gustav Pirich, Lina Goetze, and Lorenz Gschwent for excellent research assistance; all remaining mistakes are our own. We are grateful for helpful suggestions from Toke Aidt, Peter Andre, Elliott Ash, Bruno Caprettini, Milena Djourelova, Ruben Durante, Vera Eichenauer, Ruben Enikolopov, Ingar Haaland, Gaël Le Mens, Pierre-Guillaum Meon, Christopher Roth, Paul Schaudt, Marta Serra-Garcia, Johannes Wohlfahrt, Joachim Voth, as well as from participants and discussants at the Barcelona Summer Forum 2025, the Bolzano workshop on historical economics 2023, the Memories, Narratives and Beliefs Workshop in Riednau, the 4th Monash-Warwick-Zurich Text-as-Data Workshop, the EARE 2023 in Cyprus, the ifo Social Media in Economics conference (Venice), the 6th ifo Economics of Media Bias Workshop, the International Institute of Public Finance Conference, the European Public Choice Society Conference (EPCS) 2023 in Hannover, the EPCS 2025 in Riga, the Silvaplana Workshop of Political Economy, the NLP for Climate Politics Workshop, the 2023 Narrative Economics workshop in Zurich, and the Beyond Basic Questions (BBQ) Workshop in Stuttgart 2024. We also thank audiences and discussants at seminars and guest lectures at Uppsala University, at Friedrich-Alexander University Erlangen-Nuremberg,the University of Bolzano, ETH Zurich, Paris Dauphine, Hamburg University, and the University of Bern. All remaining mistakes are ours. A previous version of this paper was titled “Analyzing Climate Change Policy Narratives with the Character-Role Narrative Framework” and circulated as CESifo Working Paper No. 10429 in 2023.

†University of Bern, Wyss Academy at the University of Bern, CESifo, e-mail: *mail@kai-gehring.net*

‡University of Bern, Wyss Academy at the University of Bern, e-mail: *matteo.grigoletto@unibe.ch*

policies. *Political narratives* lead to consistently higher memory of the narrative characters, while memory of objective facts is not improved.

Keywords: Narrative economics, climate change policy, social media, virality, political economy, media economics, text-as-data, machine learning, large language models

JEL Classification: C80, D72, H10, L82, P16, Q54, Z1

1 Introduction

Politics, broadly conceived, is the social process of settling conflicts over collective decisions "through words and persuasion, and not through force" (Hannah Arendt). Such conflicts occur every day in legislatures, city councils, board meetings, workplaces, friends' circles and families. Narratives can be regarded as the tailored communication technology that actors employ to succeed in that process. *Political narratives* compress complexity about human or instrument characters into three archetypal roles called the drama triangle (hero, villain, victim) that can be processed, remembered and, crucially, shared easily. Hence, *political narratives* have an efficiency advantage in information markets when attention is limited and information processing costly. This article investigates (i.) which features enhance the virality of a political narrative and (ii.) the impact of *political narratives* on beliefs, preferences and memory.

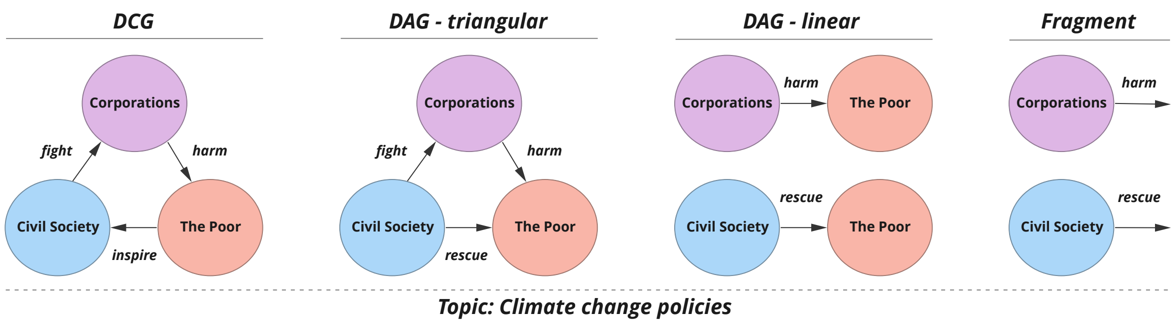
In popular science books (Harari 2014; Shiller 2020), but also increasingly in economics (e.g., Bénabou, Falk, and Tirole 2020; Bursztyn et al. 2023; Esposito et al. 2023) narratives are now generally recognized as a key communication technology that influences human preferences, beliefs, and decisions. Andre et al. (2025); Eliaz and Spiegler (2020) and Kendall and Charles (2022) started analyzing narratives systematically as causal sequences, and Barron and Fries (2023) evaluate the persuasive power of such causal narratives. However, the importance of narratives does not solely stem from their persuasive power, but as Shiller (2017) highlights, as well as from their ability to go viral. When studying virality in applications with real-world data, focusing on causal sequences has important limitations. In much of human communication – be it news, social media, or speeches – narratives are often fragmented and sometimes lack explicit causal structure, or rely on emotions, framing, tone, and dynamics between characters. To complement these existing studies, we conceptualize a broader class of *political narratives*, integrating insights from various other disciplines, as well as a measurement pipeline using large language models (LLMs).

This enables us to study the two key features of narratives: virality and persuasive power. We study virality where it is most visible: the retweet metrics on the social media platform Twitter/X. We analyze 1.15 million climate change policy tweets over the 2010 to 2021 period, encoding twenty pre-defined human and instrument characters as being in a neutral or a hero-villain-victim role. *Political narratives* are markedly more viral than otherwise similar messages, even conditional on author characteristics, text quality and emotions. Villain and hero roles produce the largest individual boost and combinations that add extra villain characters further raise virality. In complementary, pre-registered survey experiments with 3000 respondents, political narrative exposure shifts beliefs and real-money donations in the predicted direction, yet leaves stated policy positions largely unchanged. The narrative treatments also improve recall of characters, while memory for numeric facts does not improve. Hence, the power of *political narratives* both from repeated exposure to viral narratives, as well as from individual persuasiveness and improved memory.

But let's begin by better illustrating what constitutes a political narrative and how to move from concept to empirical measurement. The purpose of a political narrative is influencing perceptions, beliefs, and preferences about the characters contained in the narrative – hence the term "political'. *Political narratives* exert their influence by depicting characters in one of three archetypal roles – called the "drama triangle" (Karpman 1968) – hero, villain, or victim. Characters may be human – individuals or collective actors such as corporations, states, or movements – or instrument, denoting policies, laws, or technologies. Consider as an example, the following quote by climate activist Greta Thunberg:

*"**Global greenhouse emissions** are still on the rise, oil production is soaring and energy companies are making sky-high profits while countless people struggle to pay their bills. [...] A critical mass of people – especially younger people – are demanding change and will no longer tolerate the procrastination, denial and complacency that created this state of emergency."*

Greta Thunberg, *The New Statesman*, 19th Oct. 2022



For a given topic – here, climate-change policy (*global greenhouse emissions*) – the passage shows that naming the key characters and assigning them roles captures the essence of a political narrative. Corporations (*energy companies*), the poor (*countless people*), and civil society (*younger people, activists*) are the characters, depicted as the nodes in the graphs above. The dynamic (a)cyclical graphs (DCG/DAG) below the quote show that assigning causal arrows between characters may often be ambiguous in real texts, moreover, any causal representation presupposes that the relevant nodes have been defined and measured. By contrast, role assignment is typically clearer and can be coded directly: in this example, corporations are cast as villain, the poor as victim, and civil society as hero. This character-role coding recovers both grand narratives with multiple character-roles (left DCG/DAG graphs) and shorter fragments that are more common in natural text (right fragments of DCG/DAG graphs). Definition and measurement, therefore, reduce to specifying the topic and characters, and coding for each character whether it appears as neutral or is cast as hero, villain, or victim.

> *"A political narrative is identified by (i) its topic, (ii) its characters, and (iii) by having at least one character cast in a drama triangle role (hero, villain, or victim)."*

We provide a general pipeline, implemented in Python, to measure *political narratives* with LLMs for any topic and user-defined set of characters. The pipeline prompts an LLM to (i) detect the topic, (ii) flag the presence of specified characters, and (iii) assign roles where present, returning structured outputs that feed directly into statistical analysis. The ubiquity of the archetypal roles of the drama triangle in human texts ensures that even simple and cheap established LLMs – we use GPT-4o-mini from OpenAI – perform well even with simple one-shot prompting. Our pipeline returns a compact panel of variables – one row per tweet and columns indicating character presence and role assignment – that is straightforward to use for further analysis.

To analyze what makes *political narratives* go viral, we choose climate change policy as a topic and focus on social networks, specifically on the social network Twitter/X. Climate change policy is a suitable topic for this inquiry: its long time horizon denies participants quick feedback on outcomes, leaving narratives largely unverified in the short run and therefore dependent on their virality for influence. Because the underlying issue is characterized by deferred pay-offs and distributional conflict, *political narratives* have ample room to shift preferences and beliefs by assigning credit or blame without near-term falsification. Social media has been the focus of much of recent research in economics (see overview in Aridor et al. 2024), and has the advantage of offering clear metrics of virality. We select Twitter as a platform that became a central arena for agenda-setting and narrative contestation in Western politics (Halberstam and Knight 2016; Acemoglu, Hassan, and Tahoun 2018; Macaulay and Song 2022) over our sample period from 2010-2021. Twitter allows us to cover more than a decade of US discourse, tracking how *political narratives* evolve alongside the shifting public support patterns reported in cross-country surveys (Andre et al. 2024; Dechezleprêtre et al. 2025).

To explore our framework's capabilities, we pre-define ten characters, five human (e.g., corporations, US people) and five instrument (e.g., fossil industry, green technology, regulation) characters based on the literature, exploratory text analysis (e.g. topic models, word clouds) and our own reading of a large set of example tweets. Using climate-policy keywords following Oehl, Schaffer, and Bernauer (2017), we collect over three million English-language tweets via the Twitter API, sampling every Saturday plus one randomly chosen day per month from 2010 to 2021. We further use the metadata to select tweets originating from the US, ending with roughly 1.15 million tweets. Then we apply our pipeline, using the GPT-4o-mini model from OpenAI, to first further validate whether a tweet fits the topic. For each in-topic tweet, we detect whether each pre-defined character is present and, if so, whether it appears neutrally or in a drama triangle role: hero, villain, or victim.

We then begin by examining key text features that could be relevant for the virality and persuasive power of *political narratives*. We distinguish sentiment valence (positive–negative) and discrete emotions (e.g., joy, fear) from language metrics (e.g. length, readability). In our sample, narratives tend to employ more emotional language, with each role relying on different types of emotions. Using a standard emotion lexicon, we find that hero narratives are characterized by expressions of joy and surprise, victim narratives evoke fear, and villain narratives generally convey more anger and disgust. At the same time, we find no major differences on the language metrics. For instance, narratives are not generally easier to read or more densely written. Emotions and sentiment are one channel through which role assignment is signaled; roles can also be conveyed by explicit character labels, causal language or other cues.

We begin with descriptive results on political narrative frequency over time, which show marked shifts between 2020 and 2021. The Green Tech–Hero character-role is most frequent, followed by Fossil Industry–Villain and Corporations–Villain, while the US public appears often as both hero and victim, with notable changes over time. Beyond individual character–roles, the co-occurrence within a tweet of multiple roles reveals details about the *political narratives* over that time period. Taken together, these patterns provide an initial map of how *political narratives* – measured through the character–roles and their combinations – evolve in the climate-policy discourse.

In our main analysis, we exploit the fact that the retweet metric on Twitter constitutes a natural proxy for virality.[1] We begin by showing that the data-generating process of retweets seems to follow a power-law distribution: approximately 80% of tweets receive no retweets at all, with a rather small percentage of tweets receiving a large share of retweets. Using Poisson Pseudo-Maximum Likelihood regression models (PPML) suitable for count outcomes like retweets, we first assess the impact of simply featuring any political narrative, i.e., the tweet concerns the topic and includes at least one character cast as hero, villain, or victim. The presence of a political narrative has a positive and highly statistically significant effect on virality, which persists after controlling for author characteristics and character fixed effects. On top of that, we then control for sentiment, emotions and language metrics, evaluating whether this virality premium is solely or mostly driven by emotionality and text quality. However, the positive relationship with virality remains clearly positive and significant, highlighting that character-role combinations capture something more fundamental about the texts.

Next, we use the more detailed information from our pipeline about characters, roles and their combinations to investigate the determinants of political narrative virality. Our results indicate that both hero and villain narratives consistently boost virality compared to neutral tweets, while victim narratives do not exhibit a significant effect. In models that include all roles, villain narratives emerge as the primary driver of engagement, often amplifying the impact of hero narratives. These findings are initially derived from tweets featuring a single character-role; however, when we examine tweets with multiple roles, a clear pattern emerges: increased narrative complexity generally reduces virality, particularly when a tweet features a mix of different roles such as hero and victim or all three roles together. The only exceptions occur when a hero is paired with a villain or a victim with a villain – configurations that are more viral than a hero alone. Overall, the reinforcement of villain narratives appears to be the most influential factor in driving engagement.

Our observational analysis has several limitations. First, Twitter/X was widely used in the US across partisan lines during our sample period, but users do not systematically represent the general US population. Second, the climate change policy discussion in the US context is not identical with that in e.g. the European Union, necessitating further studies of other countries and settings. Third, future studies should explore whether similar patterns can be found for other topics. We view climate policy and Twitter/X as a tractable first setting; the measurement framework and pipeline are general and can be applied to other topics and corpora. Third, our analysis of virality

[1] We use "likes" and "replies" as another measure of narrative success. Although not equivalent to retweets and not obviously a direct proxy for virality, we observe very similar patterns. All results shown in the appendix.

is correlational: despite extensive controls and fixed effects, selection and confounding cannot be entirely ruled out. Fourth, platform ranking and the Twitter algorithm can itself shape what is seen and shared. Controlling for sentiment, emotions, and language features somehow limits the issue, but algorithmic amplification remains a residual concern that cannot be avoided when studying social networks. Fifth, automated accounts (bots) may affect diffusion dynamics; we apply standard filters for robustness tests, but acknowledge that social bots remain an endemic feature of social media.

To complement the observational analysis of virality, we conduct a series of online survey experiments studying the persuasive power of single exposure to specific *political narratives*. Specifically, we compare the effects of a treatment panel of social media posts containing a political narrative with an active control panel of comparable length and complexity that contains the same characters in a neutral role. We find that narratives shape beliefs to some degree when two hero characters are present, but much more strongly with two villain characters. *Political narratives* also significantly shift revealed preferences about a character, using GREEN TECH as a character and a pro-green technology donation as a real-stakes outcome. Single exposure to the political narrative does not significantly shift concrete policy preferences, further highlighting the importance of repeated exposure[2] ("mere exposure effect") boosted by higher virality. Finally, *political narratives* do enhance memory: participants recall posts containing narratives better. However, what they remember better are the characters (and their roles), not objective facts embedded in the narrative.

Contributions: We develop the political narrative framework as a novel approach for analyzing narratives – a narrative is defined by its topic and its characters, with at least one character cast as hero, villain, or victim. This representation complements the "narratives-as-causal-sequences" approach in economics (Eliaz and Spiegler 2020; Andre et al. 2025; Kendall and Charles 2022; Barron and Fries 2023), and also captures widely used non-sequential narrative statements, while remaining compatible with DAG/DCG causal representations. Drawing on successful applications – though mostly qualitative – in political science (Terry 1997; Jones and McBeth 2010; Jones 2014; Jiangli 2020), sociology (Polletta et al. 2011; Merry 2016; O'Brien 2018), communication studies (Anker 2005; Gomez-Zara, Boon, and Birnbaum 2018), and literary studies (Fog et al. 2010), we bring this role-based taxonomy to economics to formalize a concise and parsimonious representation of *political narratives*. The framework is topic-agnostic, making it portable across domains of interest to economists whenever a topic and a set of salient characters can be specified.

Second, we provide a Python-based pipeline for empirical measurement that takes as inputs a topic and a user-defined list of human and instrument characters and returns, for each text unit, character presence and role (hero, villain, victim, neutral). Using pre-tested prompts and batch-level API calls, the pipeline produces a structured panel – one row per document and columns for character presence and role assignment – for further statistical analysis. The design captures a broader set of narratives that appear in real text without requiring an explicit sequence (c.f. Akerlof and Snower 2016), and it aligns with DAG/DCG representations (e.g. Andre et al. 2025): any

[2] See a review on "mere exposure effect" in Montoya et al. (2017)

causal narrative estimation presupposes defined nodes, and our character–role mapping provides those nodes even when causal direction is ambiguous. Whereas prior innovative software packages like RELATIO (Ash, Gauthier, and Widmer 2024) extract subject–verb–object sequences and defers dimension reduction and aggregation of entities into characters usually to later stages, our approach moves that dimension reduction upfront by defining characters ex ante. What we define as "characters" represents a broad set of human and instrument entities, allowing researchers to adapt the scheme readily to their setting. In contrast to prior, mostly manual, approaches to encode causal narratives, our pipeline allows measuring *political narratives* at scale in large text corpora with widely available LLMs.[3] We release python code and batch scripts together with the pre-tested prompts, instructions and simple consistency rules (e.g., mutually exclusive roles per character)[4].

Third, our empirical results on the virality of *political narratives* contribute to the growing literature in economics empirically analyzing the impact of specific narratives, which we can also view through the lens of our framework. Esposito et al. (2023) demonstrate that a specific, revisionist political narrative depicting African Americans as the villain character in the US Civil War shaped political preferences and behavior. Bursztyn et al. (2023) show how exposure to specific narratives about the COVID pandemic emphasizing either China or the US democratic party as the villain changes beliefs and even high-stakes health-related behavior. Our results on virality show for a broad set of *political narratives* with the topic climate change policy which character-role combinations tend to spread more and thus could influence larger audiences. Virality has not been studied empirically in economics; the few related studies in marketing and communication find, for instance, that arousal and emotion correlate with the virality of news (Berger and Milkman 2012)[5]. Our approach allows much more systematic insights compared to examining a set of text features or different emotions by showing which character types (human vs. instrument) and which roles and role combinations are linked to higher virality on social media.

Fourth, beyond virality, our pre-registered experiments complement recent causal evidence on the persuasiveness of narratives . Andre et al. (2025) study *causal* narratives – explicit action → consequence – and show that such sequences can causally shift beliefs; Barron and Fries (2023) identify the mechanics of narrative persuasion, emphasizing sense-making and fit to facts. In contrast, we test *political* narratives defined by character–role assignments and find that single-exposure treatments systematically move beliefs and a real-stakes choice about the named character, even without specifying an explicit causal chain. Thus our evidence is complementary: it brings role-based persuasion – closer to how *political narratives* appear in practice – into a controlled setting, while the causal-sequence studies pinpoint mechanisms at a finer level of detail.

Fifth, our experiments identify a memory channel through which single-exposure to *political*

[3] Similar to other recent paper that also employ LLMs or deep-learning models in innovative ways (e.g. Ash et al. 2021; Lagakos, Michalopoulos, and Voth 2025; Voth and Yanagizawa-Drott 2025) to encode long stories and images, we demonstrate how LLMs allow us to move beyond simple metrics like emotions and study more structural features of narratives systematically.

[4] Contact the authors for an early version of the package and pipeline, which will be released after publication.

[5] Caesmann et al. (2021) study virality and persuasion, but focusing on the virality of propaganda transmitted not through media, but through public events.

narratives matters: characters are better recalled if they are cast in a drama-triangle role, whereas there is no effect on numerical facts embedded in the narratives. Compared to Graeber, Roth, and Zimmermann (2024), who compare stories to statistics and manipulate cues directly in a very controlled setting, our experiments are closer to real political content as in Barrera et al. (2020), varying role assignment while holding format and objective information constant. Decomposing the effect of different roles shows hero role assignment increases recall; adding villain roles increases recall of these characters even more strongly while crowding out the memory of hero characters. Consistent with the emerging literature on attention economics (Serra-Garcia 2025; Loewenstein and Wojtowicz 2025), scarce attention resources in settings with information overload means content needs to secure attention to be remembered. The drama-triangle roles receive attention by activating existing heuristics and mental models of the world, and villain characters who pose a potential threat receive more attention than hero characters. These findings invite future research to test in more detail aspects like cue similarity and interference in political contexts, and an investigation of how senders strategically use role assignments and role combinations to shape what is remembered.

Sixth, our results complement the growing, often survey-based, literature on the political economy of climate change. Andre et al. (2024) show that correcting misperceived norms increases support for climate change policies, and Dechezleprêtre et al. (2025) provide cross-country, experimental evidence that perceptions and ideology are more important than knowledge and understanding in shaping climate policy preferences. Djourelova et al. (2024) show that experiences of natural disasters and their media coverage affect climate change beliefs. We identify a key challenge to a constructive climate discourse by showing which *political narratives* are most likely to shape norms in real social media interactions: villain and human character messages are systematically more viral than those with instrument characters and hero roles. This creates an unfortunate incentive for both proponents and opponents of climate change policies to use blame, especially of human characters, contributing to in- vs- out-group thinking and polarization instead of focusing on constructive, mutually beneficial reforms. We then show that assigning a drama-triangle role to a character, while keeping objective embedded facts the same, can shape beliefs and memory and thus contribute to creating (or correcting) misperception. Overall our results help to better understand the reality of the climate change discourse, but also suggest that narrative-based interventions might be potentially promising compared to purely information-based approaches c.f. Haaland, Roth, and Wohlfart (2023).

Our work contributes to a large literature in media economics (see overview in Zhuravskaya, Petrova, and Enikolopov 2020) by providing a novel framework with application pipeline to analyze narratives in the media and by analyzing virality as an understudied outcome determining the effect of media on beliefs and preferences. Foundational work has examined newspapers (Gentzkow and Shapiro 2010; Djourelova, Durante, and Martin 2025), television (e.g., Ash et al. 2024a; Ash et al. 2024b; Ash and Galletta 2023; Ash and Poyker 2024; Enikolopov, Petrova, and Zhuravskaya 2011; Durante, Pinotti, and Tesei 2019; Qian and Yanagizawa-Drott 2017) and radio (e.g., Yanagizawa-Drott 2014; Adena et al. 2021). An important focus has been biases in media reporting and their sources (Durante and Knight 2012; Cage et al. 2022), strategic decisions in the timing of publications (Durante and Zhuravskaya 2018; Djourelova and Durante 2022), and the impact of competition (Cagé, Hervé, and Viaud 2020; Cagé 2020). Related studies in political economy study folklore (Michalopoulos and Xue 2021) and movies (Michalopoulos and Rauh 2024), and narratives in a historical context Cagé et al. (2023). Closely related is (Bursztyn et al. 2023), who examine how exposure to different narrative on television can have downstream effects on beliefs and behavior

Finally, we contribute to the specific literature on the economics of social media (see overview in Aridor et al. 2024) by shifting the focus from platform exposure to the structure of the content that spreads. Prior work shows that social media can shape offline outcomes, including protest participation (Enikolopov, Makarin, and Petrova 2020) and hate crimes (Müller and Schwarz 2021; Müller and Schwarz 2023). Our approach complements this evidence in two ways. First, we provide a scalable, pre-defined content measure – character–role assignments – that predicts which *political narratives* become more viral on the platform, rather than proxying by topic keywords or sentiment alone (cf. Braghieri et al. 2024). Second, we connect those content primitives to persuasion outcomes in experiments, showing that single exposure shifts beliefs and revealed preferences about the characters embedded in the narrative, while leaving policy positions largely unchanged. Taken together, the findings help reconcile why some *political narratives* persist online: human characters (especially in villain roles) carry a systematic virality premium, which increases the likelihood of repeated exposure and thereby raises the potential for downstream behavioral impact.

2 Definition and Measurement of Political Narratives

2.1 Definition

Bergstrand and Jasper (2018)

We take a definition of *political narratives* that complements the "narratives-as-causal-sequences" approach in economics while returning to the broader spirit that treats narratives as communicative devices for focusing attention, encoding roles and identities, and shaping norms and behavior (Shiller 2017; Akerlof and Snower 2016). Formally, fix a topic T and a universe of characters $K = H \cup I$, partitioned into human characters H (individuals or collective actors such as corporations, parties, states, movements) and instrument characters I (policies, laws, technologies). For any text unit (tweet, paragraph, article), let $K' \subseteq K$ be the set of characters that appear. A role-assignment function

$$r : K' \to \{\text{hero}, \text{villain}, \text{victim}, \text{neutral}\}$$

maps each appearing character to either a drama-triangle role (Karpman 1968) or neutrality. We call the triple (T, K', r) a *political narrative* if and only if at least one $k \in K'$ has $r(k) \in \{\text{hero}, \text{villain}, \text{victim}\}$; if all characters are neutral, the text is about the topic but does not constitute a political narrative in our sense. This definition intentionally accommodates fragments and non-sequential formulations (e.g., "Corporations are villains") while remaining compatible with causal or temporal representations.

The Greta Thunberg passage introduced earlier provides a concrete illustration. The characters Energy Companies, Countless People, and Younger People/Activists are the nodes (K') in the graphs shown above. The DCG/DAG panels to the left depict possible causal linkages among these nodes. In real text, however, the direction of causal arrows is often ambiguous even when role assignment is clear. Our coding therefore starts from the nodes and their roles, which can be defined and measured directly; causal arrows, when identifiable, can then be added on top of this foundation.

There are many possible ways of condensing a large set of entities into a smaller set of roles, but the drama-triangle roles are the smallest set that still preserves sufficient evaluative contrasts. It collapses many possible identities and roles into three broad archetypal roles, offering a parsimonious yet expressive partition of the identity space. Some alternative schemata like Aristotle's Poetics focus more on functional roles within a plot and with regard to story progression. Others feature larger sets of roles, ranging from 6-8 by Campbell, 7 in Propp's prominent morphology, to 12-16 archetypes according to Carl Jung. Many of these more distinct identities and roles can be meaningfully collapsed into the three drama-triangle roles (see evidence in Bergstrand and Jasper 2018).[6] They provide a recurrent role schemata that is reflected in human story telling from biblical parables and classical epics to contemporary news and social media.

Table 1 clarifies how this role-based definition relates to causal-sequence approaches and illustrates various ways of assigning roles. Let $G = (K', E)$ denote a directed (a)cyclical graph over the appearing characters, where $E \subseteq K' \times K'$ collects causal or temporal arrows. A text qualifies as a *causal* narrative when $E \neq \varnothing$; it qualifies as a *political* narrative in our sense when $\exists\, k \in K'$ with $r(k) \in \{\text{hero}, \text{villain}, \text{victim}\}$. The rows of Table 1 then parse common cases. Statements like "A carbon tax is meant to raise the price of certain goods" satisfy $E \neq \varnothing$ but leave all $r(k) = \text{neutral}$ (causal only). "The carbon tax is stupid!" assigns an evaluative role to a policy instrument ($r(\textsc{Carbon Tax}) = \text{villain}$) without specifying arrows (political only). "Price increases due to the carbon tax raise the cost of living for Americans" both links nodes ($E \neq \varnothing$) and implicitly casts Carbon Tax as a villain to US People (both causal and political). Finally, "Tariffs are beautiful" is a fragment that assigns a (positive) stance but does not specify a sequence (political only).

Crucially, the transition from neutrality to a role can be signaled in multiple ways that our framework allows: via evaluative labels and attributions (e.g., "greedy," "reckless"), via agency and responsibility in the syntax (who helps whom, who harms whom), via contrasts in evaluation across characters within the same sentence, and – when present – via the position a character occupies in an explicit causal chain. Thus emotions and sentiment can *signal* roles, but they are neither necessary nor sufficient: a text can be emotionally flat and still cast a villain (e.g., "Firm X caused

[6] Reducing roles to a simple positive–negative polarity erases a key evaluative contrast: the distinction between active agents and passive sufferers. In a sentence like "Corporations exploit Workers," both entities receive a negative valence if collapsed to "bad," yet the narrative logic depends on one being an active perpetrator and the other a passive target. Many alternative role systems explicitly include one or more victim roles (e.g. Propp and Jung).

Table 1: *Examples of Narratives*

Feature			Examples		Types of narrative	
Sequences (causal, temporal)	*Emotions*	*Character-Role(s)*	*General*	*Climate Policy*	*Causal Narratives*	*Political Narratives*
✓			Tariffs affect the terms of trade.	A carbon tax is meant to raise the price of certain goods	Yes	No
	✓		Recently, I became very curious regarding news about tariffs	Recently, I became very curious regarding news about the carbon tax	No	No
✓		✓	Tariffs on foreign competitors protect domestic producers	Price increases due to the carbon tax raise the costs of living for Americans	Yes	Yes
✓	✓	✓	Tariffs on greedy foreign competitors protect our struggling domestic producers	Price increases due to the stupid carbon tax raise the costs of living of vulnerable everyday Americans	Yes	Yes
	✓	✓	Tariffs are “beautiful”	The carbon tax is stupid!	No	Yes

the spill”), or emotionally charged without assigning any character a role (e.g., “I'm anxious about climate change”).

Narratives often appear as fragments – slogans or moral labels – rather than fully articulated stories. Formally, a *fragment* is a restriction of (T, K', r) to a smaller set of nodes $K'' \subseteq K'$ (and, where relevant, E to $K'' \times K''$). Such fragments typically sit inside grander, meta-narratives that readers implicitly know. In the Thunberg example, the grander narrative ties multiple characters across several sentences, whereas many real-world phrases will express only a piece of that structure (e.g., “ENERGY COMPANIES profit while people suffer”). Our character–role coding captures both the frequent fragments we observe in corpora (right-hand panels in the DCG/DAG illustration) and the richer configurations when multiple roles co-occur (left-hand panels).

This representation also clarifies measurement strategy and the role of dimension reduction. In tools like topic models or RELATIO (Ash, Gauthier, and Widmer 2024), the pipeline begins by extracting many surface forms (entities, subject–verb–object triples) and then asks the researcher to aggregate *ex post* to a manageable set of entities. In our notation, this is a mapping from a high-dimensional space into a lower-dimensional character set K (e.g., {*immigrants*, *migrants*, *refugees*, *asylum seekers*} $\mapsto$ IMMIGRANTS; {*carbon tax*, *emissions pricing*, *cap-and-trade*} $\mapsto$ EMISSION PRICING). We move this dimension reduction *upfront*: researchers define K ex ante, justify the choice, and then estimate $r(\cdot)$ for each $k \in K$ at the document level. This makes the object of interest – who is cast as hero, villain, victim, or neutral within topic T – transparent and portable across corpora. When desired, one can subsequently estimate E on the same K to recover causal structure; but identifying arrows always presupposes that the nodes have been defined.

Finally, the same logic is portable beyond climate-policy text. For monetary policy, central banks (FED, ECB) and their instruments (INTEREST RATES, QE) can be nodes in K and cast as heroes, villains, or victims depending on the narrative context. In immigration debates, IMMIGRANTS, NATIVE WORKERS, and POLICYMAKERS are natural human characters, while BORDER POLICY or SANCTUARY LAWS are instruments. In trade disputes, DOMESTIC PRODUCERS, FOREIGN COMPETITORS, and TARIFFS fill the same roles. Our framework keeps the unit of analysis – the character–role assignment – constant across settings, allowing researchers to compare narratives within and across topics and to layer causal arrows when the text, design, or auxiliary data support them.

2.2 Pipeline

Reproducibility and reuse. We release a Python package and example notebooks that (i) implement efficient batching logic and API calls, (ii) enforce simple consistency rules (e.g., one role per character), and (iii) produce a tidy panel with one row per document and columns for character presence and roles, ready for estimation. Our implementation is optimized for OpenAI (GPT-4o-mini) but is model-agnostic: any modern LLM that returns structured JSON can be used with minor modifications. The repository includes the final, pre-tested prompts and is available upon request; a fixed public link will be provided upon publication.

Objects and notation. Let $\mathcal{D}$ be the set of documents, $\mathcal{K}$ the set of user-defined characters for the topic, and $\mathcal{R} = \{\text{hero}, \text{villain}, \text{victim}\}$. The pipeline produces (i) a document-by-character presence matrix $\mathbf{M} = (m_{ik})_{i\in\mathcal{D},k\in\mathcal{K}}$ with $m_{ik} \in \{0,1\}$ and (ii) role indicators $\mathbf{R} = (r_{ikr})_{i\in\mathcal{D},k\in\mathcal{K},r\in\mathcal{R}}$ with $r_{ikr} \in \{0,1\}$ subject to a per-character exclusivity constraint $\sum_{r\in\mathcal{R}} r_{ikr} \leq 1$. If $m_{ik} = 1$ and $\sum_r r_{ikr} = 0$, the character is recorded as *neutral*. The stacked matrix $[\,\mathbf{M}\;\;\mathbf{R}\,]$ is the analysis-ready panel.

Five implementation steps. We summarize the workflow, visualized also in Figure 1; methodological details are deferred to the Appendix.

1. Select and define the topic. A precise topic definition anchors character selection and downstream analysis.

2. Identify the source and extract data. Common sources include digitized newspapers, social media, transcribed TV/radio/YouTube, and open-ended survey responses. Typical pre-processing includes language filtering, de-duplication, and optional geo-filtering. Full extraction details and data sources are reported in Subsection A.1 and Table A.1; geo-filtering is described in Subsection A.2.

3. Identify relevant characters. Character selection maps the topic into a manageable set of human and instrument nodes (cf. Section 2). Inputs can include literature review, exploratory tools (topic models, entity recognition, RELATIO), and domain reading. This step fixes the columns of $\mathbf{M}$ and the blocks of $\mathbf{R}$ that the model will fill.

4. Prepare the prompt(s). Prompt design specifies the mapping from raw text to $(\mathbf{M}, \mathbf{R})$. We use a two-stage structure: (i) a topic-specific relevance classifier (`irrelevant`/`assert`/`deny`/`relevant`); (ii) conditional on `relevant`, character presence and role assignment over $\mathcal{K} \times \mathcal{R}$. The package enforces basic consistency (e.g., one role per character). Final prompts are reproduced in Subsection A.3.

5. Obtain predictions and assemble outputs. We annotate documents via API (using batch submission for scale), parse JSON responses, and write a tidy panel with (i) Stage-1 flags, (ii) presence m_{ik}, and (iii) role dummies r_{ikr}. Figure 1 visualizes the classification flow. Operational annotation details are in Subsection A.3.

Quality control and validation. We recommend a light human audit and, where feasible, a small human-vs-LLM comparison. In our application, we conducted a 500-tweet MTurk exercise with two coders per tweet (28 workers in total). Our evaluation showed that agreement between GPT and humans on character presence and roles was comparable or sometimes even higher than inter-human agreement (see Appendix B).

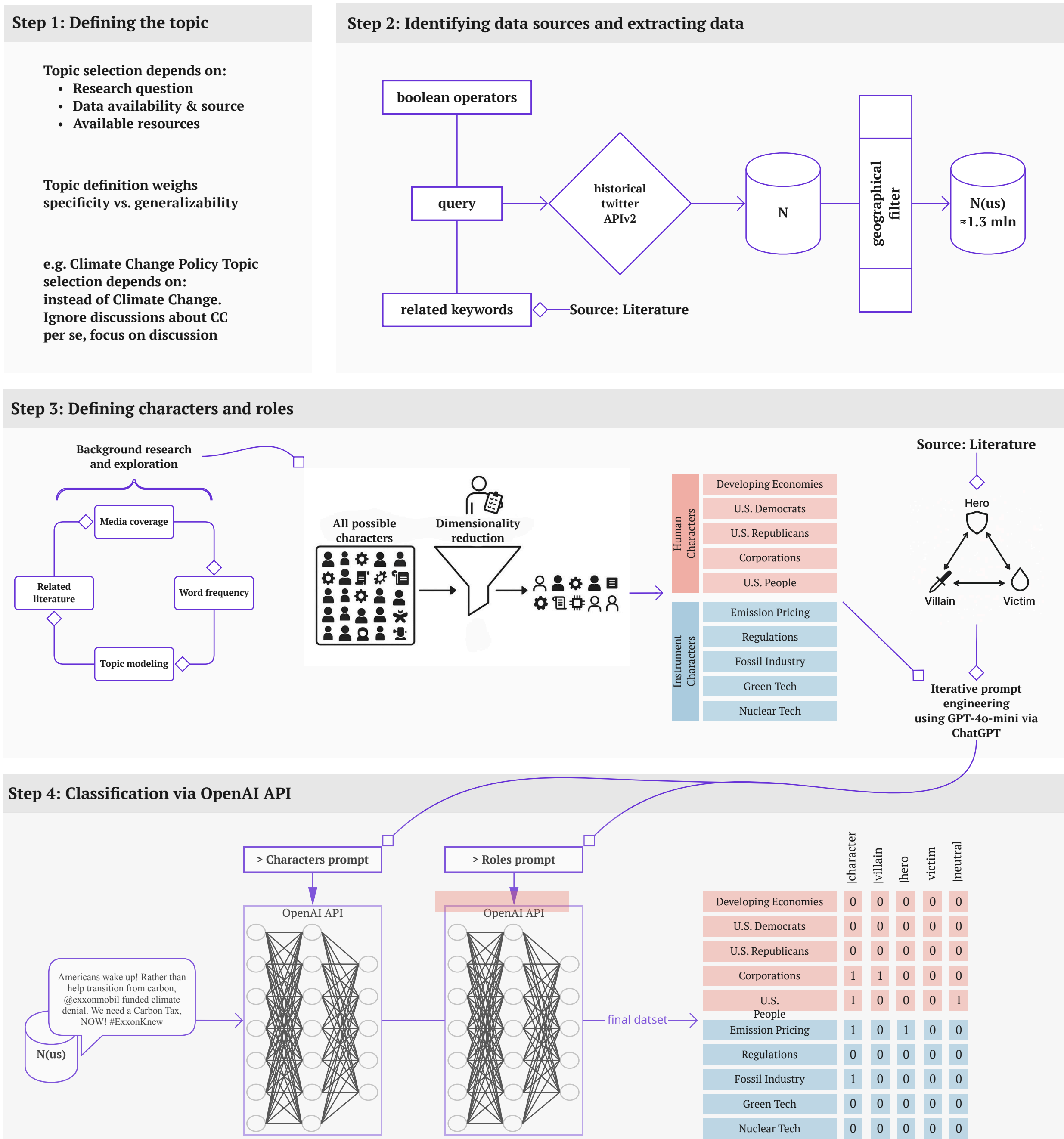
3 Data: Twitter Sample over 2010–2021

This section applies the five-step pipeline in §2.2 to our setting; methodological details are deferred to the Appendix.

1. Topic. We study *political narratives* about climate policy (not climate science) in the United States over 2010–2021.

2. Source and extraction. We query Twitter/X's historical APIv2 with a climate-policy keyword set adapted from Oehl, Schaffer, and Bernauer (2017), combining (i) one randomly selected day per month and (ii) every Saturday, to balance representativeness and tractability. We restrict to English, exclude retweets (retweet counts are used later as outcomes), and de-duplicate tweet IDs. We further validate language using `spaCy`'s language detection and normalize text (strip non-BMP unicode, standardize line breaks); details in Subsection A.1. We geo-filter tweets to the US using user profile locations and, when available, tweet geo-tags matched via `geopy`/Nominatim to OpenStreetMap

Figure 1: ***Visualization of the Classification Process***



Notes: The diagram illustrates Stage 1 (topic relevance) and Stage 2 (character presence and role assignment) for each tweet. The implementation parallelizes request preparation and uses API batch processing to scale to millions of tokens (see Subsection A.3).

polygons; see Subsection A.2. Data sources and availability are summarized in Table A.1. After cleaning and geo-filtering, the analysis corpus contains 1,151,671 tweets.

3. Characters. Guided by the relevant literature, exploratory tools (word clouds, topic models, entity recognition, RELATIO), and in particular intensive domain reading (cf. Section 2), we pre-specify ten characters: five human (Developing Economies, US Democrats, US Republicans, Corporations, US People) and five instrument (Emission Pricing, Regulations, Fossil Industry, Green Tech, Nuclear Tech).

4. Prompts and annotation. We use a two-stage prompting scheme with GPT-4o-mini via the OpenAI API. Stage 1 flags topic relevance (`irrelevant/assert/deny/relevant`). Conditional on `relevant`, Stage 2 detects character presence and assigns at most one role (hero, villain, victim) per character, recording neutral otherwise. Prompt text and operational details are in Subsection A.3; the batch annotation flow is depicted in Figure 1. We use OpenAI's Batch modality to process the corpus efficiently.

5. Outputs and analysis sample. The annotation produces a tidy panel with (i) Stage-1 flags, (ii) character presence m_{ik}, and (iii) role indicators r_{ikr}. We define tweets as `relevant` if referring to our topic and containing at least one of our pre-specified characters (neutral or in a role). This yields 309,744 `relevant` tweets. Within `relevant tweets`, a *political narrative* is defined by the existence of at least one role assignment, i.e., $\exists\,(k,r)$ such that $r_{ikr}=1$; all other `relevant tweets` contain characters featured in no role, in a neutral way.Table C.3 reports detailed descriptive statistics, with mean words $\approx$ 29 excluding hashtags/mentions and mean retweets 3.7, with heavy-tailed dispersion.

Quality control and validation. We conducted a Mechanical Turk exercise on 500 randomly sampled relevant tweets, with two independent human coders per tweet using the same role taxonomy and instructions. Setup and agreement statistics are reported in Appendix B.

4 Descriptive Evidence

4.1 Frequency of character-roles

This section maps the landscape of *political narratives* in our corpus by describing the frequency of character–role assignments for all `relevant` tweets. Using the notation introduced in Subsection 2.2, let $m_{ik} \in \{0,1\}$ indicate the presence of character $k \in \mathcal{K}$ in tweet $i \in \mathcal{D}$ and $r_{ikr} \in \{0,1\}$ the assignment of role $r \in \mathcal{R} = \{\text{hero}, \text{villain}, \text{victim}\}$. Appendix Figure C.2 details corpus composition: roughly 16% of climate-policy tweets mention no listed character, our final sample consist of the 15% feature with characters that are all only in neutral roles, and the remainder where at least one character is cast in a drama-triangle role. Table 2 summarizes character–role shares among `relevant` tweets. Panel (a) covers human characters and Panel (b) instrument characters.

Some patterns stand out. Among instrument characters, Green Tech as hero and the Fossil Industry as villain are the most frequent role assignments (both around 11–12% of all relevant tweets; Panel B). Second, among human characters, the US People is a prominent character appearing both as heroes (about 4%) and as victims (about 3%; Panel A). Corporations appear

Table 2: ***Share of Character–Roles in Relevant Tweets (United States, 2010–2021)***

Panel A: Human Characters

	Hero	Villain	Victim	Neutral	Total
Developing Economies	0.13	1.20	0.90	0.20	2.43
US Democrats	5.70	1.81	0.06	1.35	8.92
US Republicans	0.12	9.46	.	1.58	11.16
Corporations	0.98	8.14	0.06	7.87	17.05
US People	3.99	0.55	3.09	9.68	17.31

Panel B: Instrument Characters

	Hero	Villain	Victim	Neutral	Total
Emission Pricing	2.04	1.25	.	1.19	4.48
Regulations	3.48	1.36	.	3.18	8.01
Fossil Industry	0.06	11.75	0.04	1.60	13.46
Green Tech	11.52	0.84	.	3.19	15.55
Nuclear Tech	0.86	0.32	0.02	0.44	1.64

Notes: Shares are computed over the set of `relevant` tweets (those with $\sum_k m_{ik} \geq 1$). We report character–roles that appear at least 100 times over 2010–2021; excluded cells are shown as dots. "Neutral" indicates $m_{ik} = 1$ and $\sum_r r_{ikr} = 0$ for that character. Multiple character–roles may co-occur within a tweet. Appendix Table C.6 reports absolute counts (including categories excluded here).

frequently as villains (about 8%). US Republicans as villains and US Democrats as heroes appear rather frequently, showing a clear role assignment within this policy space. We will not focus on parties as characters in this paper, but defer a more detailed analysis to future work. At the other end of the distribution, some role–character pairs are very rare or non-existent in this corpus (e.g., US Republicans–Victim, Emission Pricing–Victim, Regulations–Victim, Green Tech–Victim). All character-roles appearing less than 100 times over our sample period are excluded from our further analysis. Appendix Table C.6 shows the absolute counts.

4.2 Features of Political Narratives

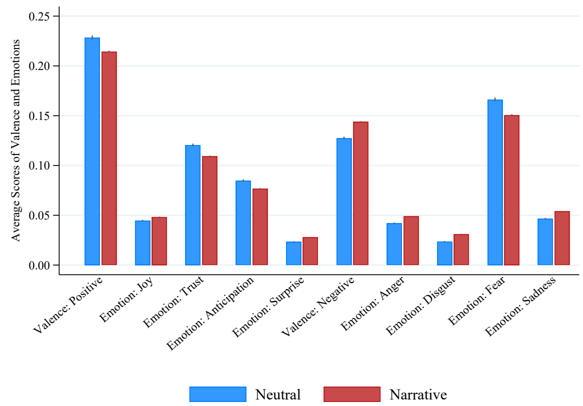
In this section, we explore the text features of political narratives. We create a variable *political narrative* which takes on the value 1 if at least one of the characters are assigned to a role, and the value 0 for the remaining tweets in `relevant` where all the characters are cast in a neutral role (for all $\sum_k m_{ik} \geq 1$). The aim is descriptive: to see whether *political narratives* differ systematically in emotional content/valence and in language metrics, characteristics that might affect processing and sharing of the content. Emotional content/valence is not the only way of assigning roles, as discussed, but it is a plausible and common way of doing so.

Figure 2 summarizes the comparison. The top row contrasts `political narrative` tweets, to tweets where all characters are featured neutrally; the bottom row breaks out roles. Emotions and valence are computed using the NRC Emotion Lexicon (Mohammad and Turney 2013) (joy, trust, anticipation, surprise, anger, disgust, fear, sadness; valence is overall positivity/negativity). Language metrics include reading ease, years of education needed, type–token ratio, lexical density, number of words, average word length, and average sentence length (standardized for readability).

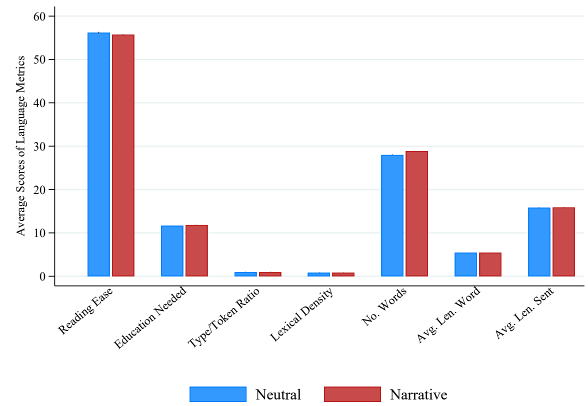
Three facts emerge. First, *political narratives* carry stronger emotional content than their neutral counterpart: when averaged across roles, they skew more negative in valence and load more on discrete emotions; this average masks clear heterogeneity by role. Second, the role breakdown is intuitive: hero narratives are the most positive; villain narratives are the most negative (with anger/disgust particularly salient); victim narratives load on fear. Third, language metrics are broadly similar across *political narratives* and neutral tweets; the main difference is length – *political narratives* are slightly longer on average, with victim narratives tending to be the longest.

Figure 2: ***Valence, Emotions, and Language Metrics of Relevant Tweets (United States, 2010–2021)***

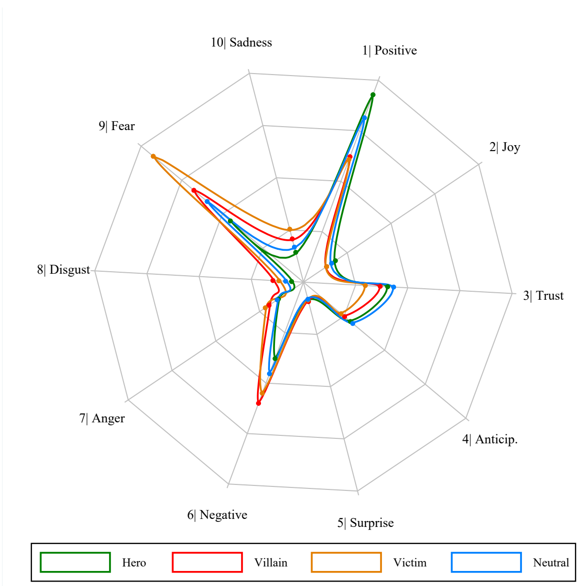
(a) Political Narrative vs. Neutral Characters: Emotions and Valence



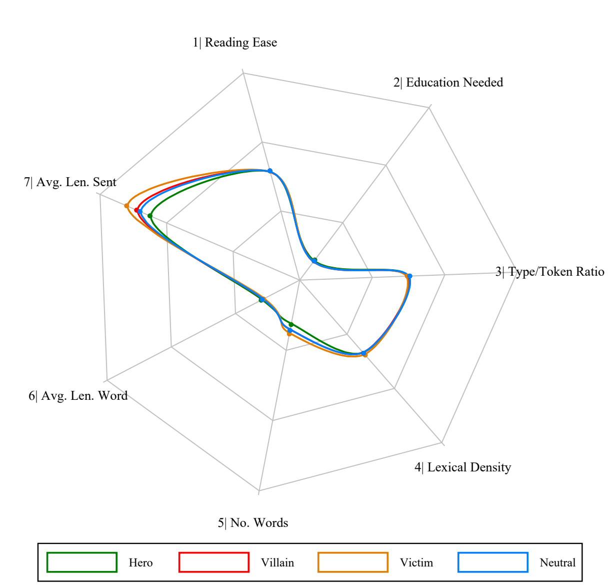
(b) Political Narrative vs. Neutral Characters: Language Metrics



(c) Across Roles: Emotions and Valence



(d) Across Roles: Language Metrics



Notes: Panels 2a–2b compare tweets with characters all in a neutral role to political narrative tweets (at least one $r_{ikr} = 1$). Panels 2c–2d break out the latter by role. Scores are standardized for readability; higher values indicate greater presence of the corresponding emotion/metric. Emotion and valence measures use the NRC Emotion Lexicon (Mohammad and Turney 2013).

4.3 Political Narratives Over Time

Compared to the pure frequency table by roles before, Figure 3 presents the share and evolution of political narratives about climate policy over time. The January 2010 to September 2021 period features significant economic and political shifts, including general events like changes in presidencies, as well as topic specific changes like the US leaving the Paris climate agreement. The figure focuses on the frequency of individual character-roles and does not yet account for their combinations within the same tweet. We organize the discussion using three ideas. A shift is any reallocation of attention across characters or across a character's roles; a reversal is a flip in the dominant role of a character or between character type; and we will speak of role entrenchment if a character's dominant roles and role hierarchy is stable over time.

From an economist's perspective, two descriptive insights stand out. First, the sharp decline in Emission Pricing–Hero, including both carbon taxes and carbon markets, suggests weaker support for price-based solutions in the public conversation, even as such policies gained widespread academic support.

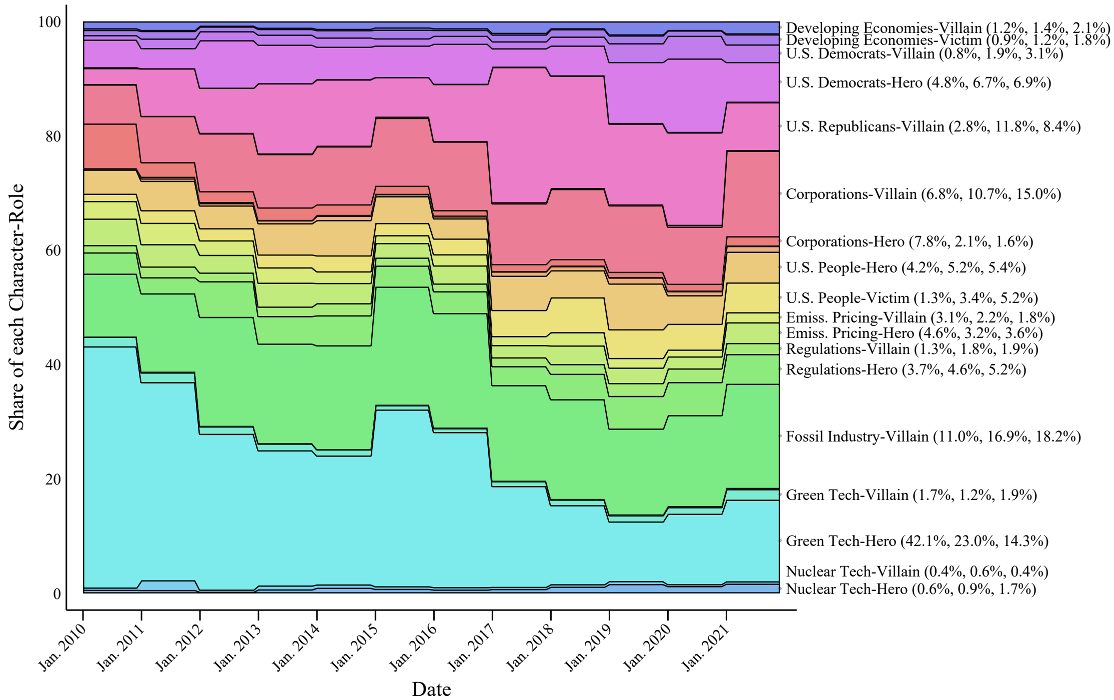
Second, the reallocation towards human characters as the expense of instrument characters can be seen as a sign of increasing populism, but also growing distrust in policy makers and the experts helping to draft policies.

We are careful not to claim causality here with regard to the causes of these change; the value is to document where attention moved in the public discourse on social media.

The composition of narratives shifts considerably over time, reflecting broader changes in climate policy discourse. Fossil Industry as a villain became increasingly prevalent, rising from 11% to 18%. US People as victims nearly triples in frequency, suggesting a growing perception that common people are suffering either from climate change or from the policies aiming to address it. A parallel trend is observed in the portrayal of Corporations as villains, which increased from 7% to 15% by 2021, reinforcing the perception of corporate actors as central contributors to climate-related challenges. One of the most striking shifts concerns the decline of the Green Tech-Hero narrative, which fell from 42% to 14%, potentially signaling a decreased optimism about technological solutions to climate change. At the same time, the increase in US Republicans framed as villains suggests a more politicized debate in recent years, reflecting heightened polarization in climate policy discussions.

Taken as a whole, these changes suggest a growing frustration with government inaction or even perceived support for fossil industries and other contributors to climate change. Beyond a declining reliance on the government as a hero, we observe a marked increase in the Corporations-Villain narrative, coupled with decreasing trust in green tech-Hero. Overall, the discussion appears to have become more politicized and polarized, reflecting a gloomier outlook on the future.

Figure 3: ***Frequency of Character-Roles over Time***



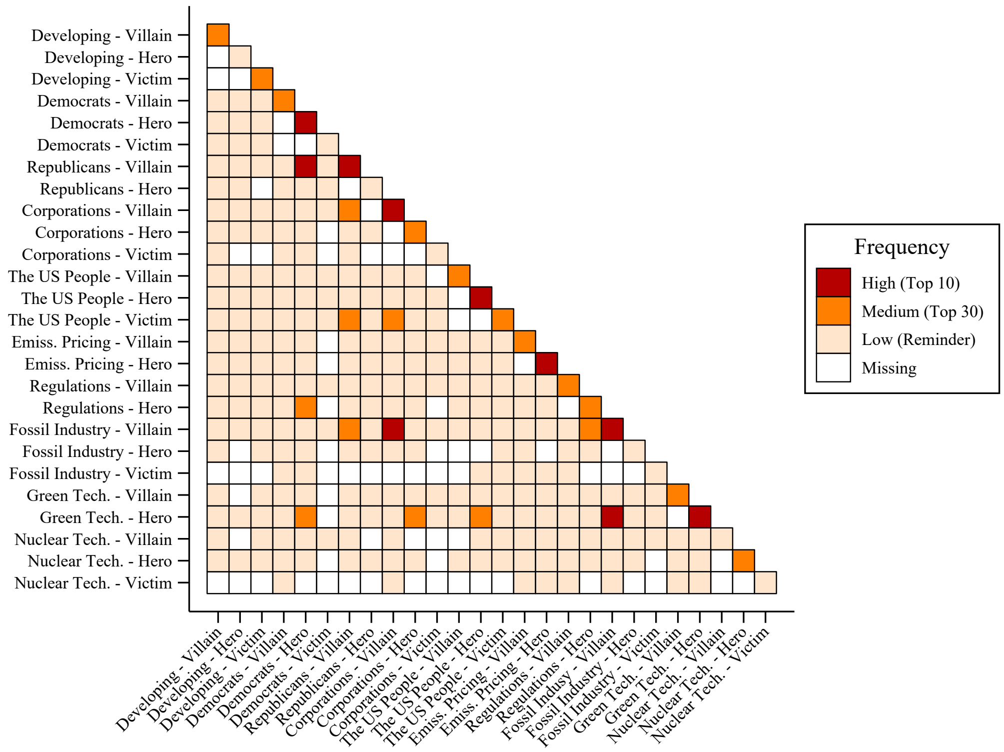
Notes: The figure plots annual shares of each character–role in the relevant-tweet dataset. For each year (2010–2021), the share of a character–role is its fraction among all identified character–role mentions that year. To maintain readability, we display labels for only 18 roles. Each label reports in order: start-year share, sample mean, end-year share. Unlabeled roles (top to bottom) are: Developing Economies-Hero, US Democrats-Victim, US Republicans-Hero, Corporations-Victim, US People-Villain, Fossil Industry-Hero, Fossil Industry-Victim, and Nuclear Tech-Victim.

4.4 Character-role combinations

While Figure 3 shows the overall frequency with which a specific character-role appears, Figure 4 depicts possible combinations of that with other character-roles, however limited to tweets with a maximum number of two character-roles, to ease the visualization. The figure shows all combinations in a matrix form, with the diagonal reporting occurrences of the character-role in simple narratives, hence without any other character-roles. For the remaining entries of the matrix, we compute for each single character-role the amount of times it appears in a 1- or 2-character-roles narrative, then compute the share of each combinations with other character-roles. One interpretation is that the lower the share displayed on the diagonal, the more complex discussions involving that character-role. Shares are not displayed numerically to avoid visual overload, but we indicate the top 5 and top 30 most frequent, with a color scheme.

For the majority of character-roles, the simple narrative form is the most common way in which they appear. In the matrix, we highlight the most frequent co-occurrences. All top 5 most frequent narratives are single character-roles narratives being US Democrats-Hero, US Republicans-Villain, Corporations-Villain, Fossil Industry-Villain and Green Tech-Hero. Nevertheless, several combinations of narratives are very common in our dataset among which US People-Victim that appears paired to Corporations-Villain. Many frequent pairs are constructed by

Figure 4: ***Absolute Frequency of Character-Roles Combinations - Tweets with One or Two Character-Roles***



Notes: The figure shows the frequency of each character–role appearing alone or in combination with another character–role, relative to all relevant tweets containing one or two character–roles. Frequencies are computed as the number of times a particular role (or pair) appears divided by the total number of tweets with one or two roles. The diagonal of the matrix shows how often each character-role appears alone in a tweet. Tweets with three or more character-roles are excluded. We do not display exact shares to avoid visual overloading. We use a color scheme to highlight the top 10 most frequent character-role combinations, the top 30 (which includes the top 10), and the remaining pairs. White indicates a pair that never appears together. The top 10 are in order: Green Tech-Hero (10.27%), US Republicans-Villain (8.95%), Fossil Industry-Villain (5.56%), US Democrats-Hero (4.21%), Corporations-Villain (3.66%), US People-Hero (3.40%), Fossil Industry-Villain + Corporations-Villain (3.27%), Green Tech-Hero + Fossil Industry-Villain (3.02%), Emission Pricing-Hero (2.40%), US Republicans-Villain + US Democrats-Hero (1.92%).

associating Green Tech-Hero to other character-roles among which US People-Hero. There are also combinations that never occur together, like Fossil Industry-Hero and Emission Pricing-Hero.

Although Twitter is dominated by short and relatively simple narratives, likely due to the platform's word limit, it is important to note that complex narratives – combining several character-roles – play also a significant role in the discourse on climate policy. Our approach is effective in capturing both simple, slogan-like narratives and more articulate narratives featuring multiple character-roles. Although our method may sacrifice some precision in capturing the exact causal links and direction of the narrative conveyed in the text, it enables the analysis of large amounts of data and allows for a nuanced depiction of the discussion.

Virality:
What Makes Narratives Go Viral, and Does it Matter?

Kai Gehring* Matteo Grigoletto**

Appendix

*University of Bern, Wyss Academy at the University of Bern, e-mail: *ka.gehring@unibe.ch*
**University of Bern, Wyss Academy at the University of Bern, e-mail: *matteo.grigoletto@unibe.ch*

Appendix

Table of Contents

A Process and methods

This appendix provides additional details on our pipeline. The aim is to facilitate its replicability and assist researchers in using our methodology for similar projects. While some details may overlap with those mentioned in the paper, this appendix provides complementing information.

A.1 Data extraction

This section outlines the data source and time frame of the extracted data. It is important to recognize potential differences in narrative structures in various sources, such as books, newspapers, and social networks. While digitized newspapers (Gehring, Adema, and Poutvaara 2022; Beach and Hanlon 2023) and social media (Cagé, Hervé, and Viaud 2020) are common sources of text data in economics, other formats, like transcribed TV, radio, YouTube broadcasts, or open-ended survey responses, also offer valuable material. Our framework is adaptable to any type of text.

In this study, we focus on English-language tweets from the United States, posted on Twitter (now X) between 2010 and 2021. At the time this project began, the historical Twitter APIv2 allowed researchers access to all tweets posted (and not deleted) since 2006. We chose the US because Twitter plays a significant role in policy discussions, and 2010 marks the point when Twitter became a mainstream platform. While the sample of US Twitter users is not fully representative, it offers a unique opportunity to observe the creation and spread of narratives over time and across different regions.

Keywords and query

We indicate here the keywords and rules used for the query of Twitter historical APIv2. Consider the following conditions:

1. The tweet includes at least one of the following terms: 'climate change', 'global warming', 'renewable energy', 'energy policy', 'emission', 'certificate trading', 'green certificate', 'white certificate', 'combined heat', 'power solution', 'energy solution', 'CO2', 'energy efficiency', 'energy saving', 'solar power', 'solar energy', 'wind power', 'wind energy', '*renewable energies*', '*energy policies*', '*ipcc*', '*green growth*', '*green-growth*', '*green wash*', '*green-wash*', '*climate strike*', '*climate action*', '*strike 4 climate*', '*strike for climate*'.
2. The tweet includes at least one of the following terms: 'climate', 'global warming', 'greenhouse' AND at least one of the following terms: 'refining', 'feed-in', 'cogeneration', 'extraction', 'exploitation', 'geotherm', 'hydro', 'agriculture', 'waste management', 'forest', 'wood', '*problem*', '*issue*', '*effect*', '*gas*', '*degrowth*', '*de-growth*', '*fridaysforfuture*', '*fridays4future*', '*scientistsforfuture*', '*scientists4future*'.
3. The tweet includes at least one of the following terms: 'climatechange', 'globalwarming', 'renewableenergy', 'renewableenergies', 'energypolicy', 'energypolicies', 'greencertificate', 'whitecertificate', 'combinedheat', 'powersolution', 'energysolution', 'energyefficiency', 'energysaving', 'solarpower', 'solarenergy', 'windpower', 'windenergy', 'greengrowth', 'greenwash', 'climatestrike', 'climateaction', 'strike4climate', 'strikeforclimate'.

4. The tweet includes term 'carbon' AND Tweet does NOT include any of the following terms: 'bicycle', 'bike', 'copy', 'fiber', 'rims', 'altered', 'fork', 'frame', 'dating', 'tacos'.

A tweet is part of our sample if any of the above conditions applies. In addition, a tweet is part of our sample if its text also satisfies all of the following:

- (a) The tweet does not contain an URL address.
- (b) The tweet's content is in English language.
- (c) The tweet is not a retweet.

The following are the changes adopted in deviation from keywords and rules proposed by Oehl, Schaffer, and Bernauer (2017), the paper of reference for us to define our query:

1. In Oehl, Schaffer, and Bernauer (2017) any keyword needs to appear in combination with at least one among: 'climate', 'global warming', 'greenhouse'. We do not adopt the condition as baseline but we use it for those words that refer to climate change in a more lose way (see condition No. 2 above).
2. We use some terms that are not present in Oehl, Schaffer, and Bernauer (2017): 'ipcc', 'climate change', 'energy policies', 'renewable energies', 'green growth', 'green-growth', 'green wash', 'green-wash', 'climate strike', 'climate action', 'strike 4 climate', 'strike for climate', 'problem', 'issue', 'effect', 'gas', 'degrowth', 'de-growth', 'fridaysforfutrue', 'fridays4future', 'scientistsforfuture', 'scientists4future' (see words in *italics* in conditions 1 and 2).
3. We use all multi-word expressions in condition 1 (e.g. 'energy policies') also as hashtags (see condition 3).
4. We use an exclusion restriction tailored towards tweets, because we realized there was a consistent pattern of false positive cases with the word 'carbon' (see condition 4).

Extracted data

In this analysis, we use data extracted via the Historical Twitter API in two distinct ways. First, we collected tweets from randomly selected days within the time frame of interest, which we refer to as the 'random days' dataset. Second, we gathered tweets from every Saturday within the same period, which we call the 'every Saturday' dataset. These two datasets were then combined into a final dataset. The random days dataset aims to provide a representative sample of tweets across the entire period. The inclusion of tweets from every Saturday helps to capture data that is less likely to be influenced by specific events, unless those events are cyclical and consistently occur on Saturdays.

Random days dataset

We collect tweets extracted from a set of randomly selected days in the period 2010-2021. We use the calendar option of the online random number generator random.org to randomly select a day within each month of this time period. The extraction was done on 7^{th} and 8^{th} February 2022. The selected days are the following:

2010-01-28, 2010-02-14, 2010-03-13, 2010-04-30, 2010-05-25, 2010-06-30, 2010-07-07, 2010-08-04, 2010-09-14, 2010-10-02, 2010-11-06, 2010-12-14, 2011-01-14, 2011-02-09, 2011-03-01, 2011-04-10, 2011-05-13, 2011-06-19, 2011-07-22, 2011-08-15, 2011-09-08, 2011-10-23, 2011-11-21, 2011-12-17, 2012-01-31, 2012-02-27, 2012-03-26, 2012-04-04, 2012-05-26, 2012-06-18, 2012-07-10, 2012-08-18, 2012-09-20, 2012-10-22, 2012-11-01, 2012-12-03, 2013-01-15, 2013-02-12, 2013-03-27, 2013-04-25, 2013-05-05, 2013-06-18, 2013-07-19, 2013-08-08, 2013-09-25, 2013-10-11, 2013-11-06, 2013-12-01, 2014-01-24, 2014-02-13, 2014-03-04, 2014-04-30, 2014-05-16, 2014-06-23, 2014-07-12, 2014-08-21, 2014-09-26, 2014-10-24, 2014-11-05, 2014-12-06, 2015-01-26, 2015-02-21, 2015-03-20, 2015-04-24, 2015-05-06, 2015-06-09, 2015-07-23, 2015-08-20, 2015-09-15, 2015-10-15, 2015-11-11, 2015-12-21, 2016-01-11, 2016-02-05, 2016-03-22, 2016-04-02, 2016-05-01, 2016-06-19, 2016-07-01, 2016-08-31, 2016-09-09, 2016-10-13, 2016-11-14, 2016-12-22, 2017-01-01, 2017-02-12, 2017-03-25, 2017-04-04, 2017-05-07, 2017-06-05, 2017-07-11, 2017-08-27, 2017-09-14, 2017-10-21, 2017-11-09, 2017-12-21, 2018-01-09, 2018-02-09, 2018-03-30, 2018-04-06, 2018-05-08, 2018-06-05, 2018-07-06, 2018-08-14, 2018-09-16, 2018-10-22, 2018-11-12, 2018-12-15, 2019-01-04, 2019-02-14, 2019-03-15, 2019-04-19, 2019-05-17, 2019-06-21, 2019-07-22, 2019-08-30, 2019-09-19, 2019-10-01, 2019-11-01, 2019-12-01, 2020-01-12, 2020-02-12, 2020-03-22, 2020-04-16, 2020-05-08, 2020-06-22, 2020-07-17, 2020-08-17, 2020-09-26, 2020-10-08, 2020-11-07, 2020-12-18, 2021-01-23, 2021-02-25, 2021-03-20, 2021-04-05, 2021-05-23, 2021-06-12, 2021-07-11, 2021-08-30, 2021-09-25, 2021-10-10, 2021-11-11, 2021-12-30.

Every Saturday dataset

We collect a large sample of tweets extracted along the same period of analysis 2010-2021. We collect tweets from every Saturday of every week between January 2010 and December 2021. The extraction was done between 4^{th} and 7^{th} December 2022.

Data managing

We compute a number of steps to clean and organize data after the extraction. We describe these steps in the following points:

1. Despite setting API's filter, some non-English tweets were captured and we had to clean them using langdetect, a python port of the language-detection library in Java. At the time of writing, langdetect is also available as an extension in spaCy.

2. The text of a single tweet might satisfy more than one condition of our query, hence representing a potential duplicate in the extracted data. Each tweet is associated to a uniquely identifying ID that we use to drop potential duplicates.

3. Before labeling, we clean the tweets of emojis and any other unicode objects that have a UTF-8 code larger than three bits. We also replace line-breaks in the text with simple spaces.

The random days dataset - after the cleaning and wrangling - comprises 1,070,702 tweets. The every Saturday dataset - after the cleaning and managing - consists of 3,279,730.

A.2 Data geo-localization

The tweets in our datasets were posted by users from around the world. Since our interest in this paper focuses on discussions in the United States, we filter the tweets to include only those originating from the US before proceeding with the annotation. In the following, we provide some indications on localization of tweets.

It is important to notice that tweets do not inherently come with localization information and this needs to be retrieved by the researchers, if possible. Many authors using tweets in their analysis developed their own methods to localize tweets (Kirilenko and Stepchenkova 2014; Baylis 2020). We build on previous work and structure our own method that exploits different 'fields' of information provided by the APIv2.

There are two main sources of geographical information available through the Historical API. Among the available user fields, there is one called 'location'. This can be filled in two ways. One method is directly by the user who decides to indicate her location when creating the profile. Another is by the Twitter API algorithm itself, which detects the location if it has been mentioned in the text of the user's self-description. For example, if a user describes herself as 'I am a PhD student based in Zurich', the API would provide Zurich as the user's location. Additionally, among the available tweet fields, there is one called 'geo'. This indicates the location of the tweet if the tweet has been geo-tagged somewhere. In fact, the Twitter application allows users to tag a tweet with a specific location at the time of posting. This simply involves indicating a place to which the user wants to 'tag' the post.

We decide to prioritize the information about the user and hence assign to each tweet the location of the user that posted it. This is because only a minority of tweets come with 'geotag' information. This might raise the suspect that people tag their tweets only during particular events - such as holidays or work trips - which do not truly represent their environment/location. Only in cases where the user's location information is unavailable do we locate a tweet according to the geo-tag assigned to it, if any. Consequently, our localization pipeline consists of the following steps, which apply to the analysis dataset:

1. We collect all available locations relative to users' profiles from the two datasets 'random days' and 'every Saturday'.

2. We use the geopy implementation of Nominatim's API which exploits OpenStreetMap data. We query the API inputting the location in string format and obtain its geographical coordinates.

3. Once each string location is associated to a set of coordinates we merge the locations back to the datasets. The merge is done with the formula 'many to one' so that users with the same location are associated to the same set of coordinates.

4. We repeat step 1, 2 and 3 for those tweets that could not be located by the description location of the users posting them but that present a geo-tag.

5. For all tweets that could be located (either via description or via geo-tag location) we intersect their coordinates with a shapefile of the United States borders and keep only those tweets located within the country.

6. The two concatenated dataset count a total of 4,236,799 tweets, after dropping potential duplicates. Once filtered for location, keeping only tweets originating from the US, the total amount is 1,151,693 tweets.

Some important notes on the Nominatim API. First, the API algorithm returns the centroid coordinates of the location, hence when searching for e.g. 'Florida' it would return the coordinates of the centroid of the state of Florida. Second, when the string location is not clear the API returns 'NaN' output. Third, in most of the cases in which the location string is composed of two or more locations - e.g. 'Florida and NY' - the API returns either one of the two or 'NaN' output. This is generally hard to predict, but multiple locations are a minority of the cases in our data. Lastly, the API is not case-sensible hence e.g. 'New York City' and 'new york city' would provide the same coordinates.

A.3 OpenAI API annotation

This section provides a guideline for the process used to annotate data via the OpenAI API. The following steps summarize the procedure with key details about the data involved in each phase.

Setting up the OpenAI API

The first step is to create a project on the OpenAI API platform. Ideally, this should be done using a business account to ensure maximum data privacy. Upon project creation, an API Key is generated, which is required for all queries and operations using the OpenAI API. The API Key can be found in the user's profile under the API Key Dashboard.

Data organization

Three main datasets were used in this process. These datasets are stored in the folder '*output - data - processed_tweets - aggregated_data*' under the names *analysis_tweets_geolocalized_usa*, *tweets_visitrump_analysis*, and *tweets_visibiden_analysis*. The first dataset includes tweets collected from random days - one per month - and every Saturday from January 2010 to December 2021. These tweets were selected using specific keywords, as detailed in the corresponding Appendix of this paper. The second and third datasets contain tweets from users active in spreading climate policy narratives, particularly during the Trump and Biden elections. They include all relevant tweets from these users within the year surrounding each election, filtered by the keywords of interest used also to collect tweets for the previous dataset described above. The three datasets were concatenated, resulting in a final dataset containing 1.15 million tweets.

Annotation modality

The final concatenated dataset represent the final data used in this annotation process. For each tweet we query the OpenAI API, prompting the selected mode (more about this below) to annotated the tweet according to our instructions. The annotation happens in two separate stages. In the first stage, we prompt the selected model to categorize the tweet's relevance to the climate change policy discussion in the US. In particular we want to know the following:

- Irrelevant: When the tweet is not really about climate change. E.g. '*The political climate is getting very day more heated!!*'
- Assert: When the tweet is about climate change but it is limited to asserting the existence of the issue, without touching onto any adaptation or response policy or action. E.g. '*#Climatechange is the single most important issue we are facing, wake up!*'
- Deny: When the tweet is about climate change but it is limited to denying its existence or proposing sarcastic and/or skeptical view on the extent of the problem. E.g. '*""global warming"" when one needs it?? It's cold outside, climate is NOT changing.*'
- Relevant: When the tweet is about climate change and it discusses related policies and issues. E.g. '*Not recognizing climate change is an issue is just insane, we need to start supporting policies that actually make a change like the carbon tax.*'

In the second stage, if and only if a tweet is found to be 'relevant' at stage 1 - thus it is about climate policy and action - then the tweet goes through stage 2. Stage 2 consists in the individuation of characters and the classification of their role in the tweet. In particular, we query the model to find the following characters:

- Developing Countries and Emerging Economies: emerging and developing economies, poorer countries, or nations part of the BRICS group -Brazil, Russia, India, China, South Africa-, as well as any related government institutions, representatives, or citizens associated with these countries and regions
- US Democrat: politicians, members, and public figures associated with the US Democratic Party. This includes prominent individuals such as Joe Biden, Nancy Pelosi, Alexandria Ocasio-Cortez, Bernie Sanders, Barack Obama, and others who represent ideals and policies of the Democratic party
- US Republicans: politicians, members, and public figures associated with the US Republican Party. This includes prominent individuals such as Trump, Mitch McConnell, Ted Cruz, Ron DeSantis, and others who represent ideals and policies of the Republican Party
- Corporations and Industry: large corporations, small and medium businesses, banks, and other private sector entities. This includes CEOs and leadership figures like Elon Musk and Jeff Bezos, representatives from industries such as energy, technology, finance, and manufacturing, as well as industry lobbying groups, corporate interests, and small or local business owners
- The People of the US: the collective citizens, voters, workers, youth, and grassroots movements of the United States, often portrayed in contrast to political elites or corporate interests. This also includes references to the 'average American,' and general terms like the public, society, community action, and any collective expression of public will or activism, including movements like FridaysForFuture, Sunrise Movement, and Extinction Rebellion

- Emission Pricing Tools: any market-based instruments and schemes designed to price carbon emissions and incentivize reductions. This includes tools such as carbon taxes, cap and trade systems, carbon pricing, emission trading, carbon markets, pollution credits, carbon credits, carbon fees, and carbon dividends. These tools are part of the broader market-led response to addressing climate change
- Banning or Regulation Policies: government actions, policies and movements aimed at combating climate change through the banning, phasing out, or strict regulation of specific products or industries. This includes efforts such as banning fracking, phasing out fossil fuels, banning single-use plastics, and other regulatory measures designed to reduce environmental impact and promote sustainability. It also encompasses movements that challenge economic growth models or advocate for systemic changes, such as de-growth movements and anti-capitalist environmental initiatives)
- Fossil Fuels: any explicit reference to fossil fuels, including terms like coal, oil, natural gas, and related critical labels such as 'dirty energy.' This encompasses all forms of energy derived from fossil sources, as well as the technologies and infrastructure that rely on them, such as combustion engines, power plants, and industrial machinery
- Green Technologies: technologies developed as a response to climate change or aimed at phasing out fossil fuels. This includes wind energy, solar energy, electric vehicles, hydrogen power, battery storage, geothermal energy, and other renewable or low-carbon technologies excluding though nuclear energy
- Nuclear Energy: all forms of nuclear energy, including both fusion and fission technologies. This encompasses nuclear power plants, nuclear reactors, nuclear fusion research, and related technologies used for energy production

For each character, the model determined if the character was present and whether they played the role of *hero*, *villain*, *victim*, or *neutral* (if no clear role applied). Tweets not deemed "relevant" in stage 1 were classified simply as either 'irrelevant', 'assert', or 'deny' (see above for explanation).

For both stages, the prompts were generated using OpenAI's ChatGPT interface. One of the authors queried the GPT-4o model in ChatGPT, providing an explanation of the task and asking the model to suggest the optimal prompt for instructing itself. Since the GPT-4o model is the same used via the API, this method ensured efficient and effective prompt construction. The final prompts used for the annotation are provided below:

Stage 1 prompt:

You are an average US citizen. The user will provide the content of a tweet posted from the US. Your task is to analyze the tweet within the context of US political discourse, particularly in relation to climate change. Respond in JSON format. 1. Relevance Check: Analyze the tweet in the context of US climate change discussion and determine its relevance. Provide one of the following values: - 0 (irrelevant): If the tweet does not discuss climate change in a meaningful way. For example, if it only includes a hashtag (like #climatechange) or a passing reference but *does not engage in any discussion about climate change or related policies, it should be considered irrelevant. - 1 (assert): If the tweet asserts the existence of climate change but does not engage with specific policies or actions related to it. This includes tweets that acknowledge climate change as an issue without going deeper into details. - 2 (deny): If the tweet denies the existence or severity of man-made climate change, referring to it as a hoax, scam, or fraud, or using sarcasm or language that undermines the reality of climate change. - 3 (relevant): If the tweet discusses climate change or related policies in a substantive way. This includes any tweet that debates, critiques, or supports policies or actions related to climate change, as well as conversations on how to combat or adapt to climate change. Respond in JSON format, returning the value in the key 'r'.*

Stage 2 prompt:

You are an average US citizen. The user will provide the content of a tweet posted from the US between 2010 and 2021. Your task is to analyze it within the context of US political discourse, particularly in relation to climate change and related policies. Respond in JSON format.1. Character Analysis: Identify whether the tweet mentions specific characters. For each character mentioned, assess their contextual role using the following scale: - Villain (1): The character is portrayed as contributing to problems, opposing positive change, negatively or engaging in harmful actions related to climate change. Look for language that blames, criticizes, or attributes a negative impact. - Hero (2): The character is portrayed as leading efforts to combat climate change, promoting environmental policies, positively or acting in a morally commendable way. Look for praise, leadership roles, or proactive efforts. - Victim (3): The character is portrayed as being unfairly attacked, facing challenges, or suffering due to external factors. Look for language that depicts them as unjustly targeted, enduring consequences, suffering or being the victim. - No role (4): Choose this option if the tweet mentions the character but does not clearly assign one of the above described roles, or if the context is ambiguous or neutral.2. Character Definitions: Evaluate these characters in the context of the tweet. - For Developing Countries and Emerging Economies (emerging and developing economies, poorer countries, or nations part of the BRICS group -Brazil, Russia, India, China, South Africa-, as well as any related government institutions, representatives, or citizens associated with these countries and regions), provide the assessment in the key 'a': - 0: No mention of the character. - 1: Villain. - 2: Hero. - 3: Victim. - 4: None of the roles applies. - For The US Democrats (politicians, members, and public figures associated with the US Democratic Party. This includes prominent individuals such as Joe Biden, Nancy Pelosi, Alexandria Ocasio-Cortez, Bernie Sanders, Barack Obama, and others who represent ideals and policies of the Democratic party), provide the assessment in the key 'b': - 0: No mention of the character. - 1: Villain. - 2: Hero. - 3: Victim. - 4: None of the roles applies. - For The US Republicans (politicians, members, and public figures associated with the US Republican Party. This includes prominent individuals such as Trump, Mitch McConnell, Ted Cruz, Ron DeSantis, and others who represent ideals and policies of the Republican Party), provide the assessment in the key 'c': - 0: No mention of the character. - 1: Villain. - 2: Hero. - 3: Victim. - 4: None of the roles applies. - For Corporations and Industry (large corporations, small and medium *businesses, banks, and other private sector entities. This includes CEOs and leadership figures like Elon Musk and Jeff Bezos, representatives from industries such as energy, technology, finance, and manufacturing, as well as industry lobbying groups, corporate interests, and small or local business owners), provide the assessment in the key ‘d‘: - 0: No mention of the character. - 1: Villain. - 2: Hero. - 3: Victim. - 4: None of the roles applies. - For The People of the US (the collective citizens, voters, workers, youth, and grassroots movements of the United States, often portrayed in contrast to political elites or corporate interests. This also includes references to the ’average American,’ and general terms like the public, society, community action, and any collective expression of public will or activism, including movements like FridaysForFuture, Sunrise Movement, and Extinction Rebellion), provide the assessment in the key ‘e‘: - 0: No mention of the character. - 1: Villain. - 2: Hero. - 3: Victim. - 4: None of the roles applies. - For Emission Pricing Tools (any market-based instruments and schemes designed to price carbon emissions and incentivize reductions. This includes tools such as carbon taxes, cap and trade systems, carbon pricing, emission trading, carbon markets, pollution credits, carbon credits, carbon fees, and carbon dividends. These tools are part of the broader market-led response to addressing climate change), provide the assessment in the key ‘f‘: - 0: No mention of the character. - 1: Villain. - 2: Hero. - 3: Victim. - 4: None of the roles applies. - For Banning or Regulation Policies (government actions, policies and movements aimed at combating climate change through the banning, phasing out, or strict regulation of specific products or industries. This includes efforts such as banning fracking, phasing out fossil fuels, banning single-use plastics, and other regulatory measures designed to reduce environmental impact and promote sustainability. It also encompasses movements that challenge economic growth models or advocate for systemic changes, such as de-growth movements and anti-capitalist environmental initiatives), provide the assessment in the key ‘g‘: - 0: No mention of the character. - 1: Villain. - 2: Hero. - 3: Victim. - 4: None of the roles applies. - For Fossil Fuels (any explicit reference to fossil fuels, including terms like coal, oil, natural gas, and related critical labels such as ’dirty energy.’ This encompasses all forms of energy derived from fossil sources, as well as the technologies and infrastructure that rely on them, such as combustion engines, power plants, and industrial machinery.), provide the assessment in the key ‘h‘: - 0: No mention of the character. - 1: Villain. - 2: Hero. - 3: Victim. - 4: None of the roles applies. - For Green Technologies (technologies developed as a response to climate change or aimed at phasing out fossil fuels. This includes wind energy, solar energy, electric vehicles, hydrogen power, battery storage, geothermal energy, and other renewable or low-carbon technologies excluding though nuclear energy), provide the assessment in the key ‘i‘: - 0: No mention of the character. - 1: Villain. - 2: Hero. - 3: Victim. - 4: None of the roles applies. - For Nuclear Energy (all forms of nuclear energy, including both fusion and fission technologies. This encompasses nuclear power plants, nuclear reactors, nuclear fusion research, and related technologies used for energy production), provide the assessment in the key ‘j‘: - 0: No mention of the character. - 1: Villain. - 2: Hero. - 3: Victim. - 4: None of the roles applies.3. Final Output: Respond with a JSON format containing the following keys: - ‘a‘: (0-4 as defined in step 2). - ‘b‘:*

(0-4 as defined in step 2). - ‘c‘: (0-4 as defined in step 2). - ‘d‘: (0-4 as defined in step 2). - ‘e‘: (0-4 as defined in step 2). - ‘f‘: (0-4 as defined in step 2). - ‘g‘: (0-4 as defined in step 2). - ‘h‘: (0-4 as defined in step 2). - ‘i‘: (0-4 as defined in step 2). - ‘j‘: (0-4 as defined in step 2).

OpenAI API Batch Modality

The final dataset used for annotation was large, and annotating it using the standard OpenAI API endpoints would have been too time-consuming. To speed up the process, we used the Batch API, which allows users to upload a large number of requests at once. OpenAI processes these requests within 24 hours, optimizing for times of lower traffic.

The size of each batch depends on the prompt and input, and more details can be found on the Batch API page. In our case, we uploaded 25,000 tweets per batch, resulting in 61 chunks for the entire dataset. Users can monitor the status of each batch on their Dashboard. Our entire dataset was processed in about a week, with a total cost of approximately 2,100 USD.

Lastly, regarding data retention: OpenAI does not use uploaded data for model training and retains it only for legal reasons, currently for one month. We recommend deleting files created via the Batch API through the Dashboard after processing.

A.4 Requirements and Sources

Table A.1: *Data Sources*

Data	Source	Download Date	Availability
Twitter Data			
Model-tweets dataset	Twitter Historical APIv2	7^{th} - 8^{th} Jan. 2022	Cannot be shared
Analysis-tweets dataset	Twitter Historical APIv2	4^{th} - 7^{th} Dec. 2022	Cannot be shared
GIS Data			
US Shapefile (v4.1)	GADM website	4^{th} Oct. 2022	Can be shared
Election Data			
Presidential election dataset	FiveThirtyEight repository	17^{th} Nov. 2022	Can be shared

Notes: The table reports a description of the sources of the data used in our analysis and their respective availability. In Section **??** we describe the process of data preparation.

B Comparison of Narratives' Classification: GPT vs Human Coders

Artificial Intelligence technology has undertaken a revolution in recent years, influencing many human activities and tasks. Among these, research is widely changing, especially when the tasks and goals include the interpretation, generation, and understanding of human language. Models like GPT are increasingly more used to classify, summarize, and extract meaning from text. These tools offer researchers a fast and scalable way to analyze large volumes of language data.

In our study, we apply GPT to a complex task: understanding and classifying *political narratives* shared by users on Twitter. Despite the drama triangle (Karpman 1968) being a widely adopted model of storytelling, deeply rooted in human communication, narratives may still leave some space for interpretation. In comparison, other NLP tasks tend to rely more on a clear "ground truth". For example, a model may classify whether a review is positive or negative, or whether a message contains offensive language. In the case of narratives, even trained Human Coders can disagree on whether a particular text expresses a given narrative (see, for example, the exploration in Gehring, Adema, and Poutvaara (2022)).

This appendix compares GPT's classifications to those of Human Coders on the same tweets. We treat GPT's output as reflecting an "average representative" Human Coder. Rather than a validation exercise, this comparison explores how closely GPT's interpretations align with those of human annotators. Below, we describe the method and present results at two levels: the character-role level and the tweet level.

Method

We started this comparison exercise by hiring workers from Amazon MTurk. To guarantee high-quality human coding, we designed a qualification task to select attentive workers. The task required participants to read the instructions for our study and answer four comprehension questions to assess their understanding. Only those who correctly answered all four questions were invited to proceed to the actual coding task. Additionally, we included a Captcha verification step to deter potential bots.

The aim of the exercise was the classification of 500 tweets, randomly selected among the tweets identified by GPT as containing at least one characters (*relevant tweets*). We structured the MTurk assignment so that each tweet was classified by two Human Coders, allowing us to compute measures of inter-coder reliability among them. In total, 80 workers successfully completed the qualification test and were invited to participate. Of these, 28 workers actually classified tweets. Workers were free to decide how many tweets they would classify. Some coded as few as one tweet, while others classified up to 130. On average, workers classified 36 tweets each. We kept the task open until we reached the objective of each of the 500 tweets being coded by two different workers.

The classification task was divided into two phases, matching as closely as possible the language and structure of the prompt used for GPT's task. For each character, the Human Coder was first asked whether the character was present in the tweet. For characters identified as present, coders then indicated whether the character was depicted as a hero, villain, victim, or none of these roles. Finally, we used these classifications to compare Human Coders to each other and to GPT. Below, we present the results.

Comparison at the Character Level

In the first part of this analysis, we compare classifications at the character level. This reflects the structure of the classification task, where Human Coders were asked, for each tweet, to evaluate each of the ten characters individually – first judging whether the character was present, then assigning a role if applicable. To mirror this structure, we organize the dataset so that each line corresponds to one character within a tweet. As a result, for each of the 500 tweets in the comparison exercise, the dataset contains ten lines, one for each character.

As a first step, we compare the overall agreement between GPT and Human Coders, and between pairs of Human Coders. Figure B.1 summarizes the results. The figure shows the share of tweets where GPT and the Human Coder agreed on the presence of the character (blue) and on the presence of the character-role (red), shown in the first two columns from the left. Importantly, GPT is not compared to the same Human Coder across all tweets but to whichever coder classified each specific tweet. The next two columns report the same agreement rates, but for pairs of Human Coders who coded the same tweet. In all comparisons, agreement includes negative agreement, thus cases where both coders (or GPT and a Human Coder) agreed that the character or character-role was absent.

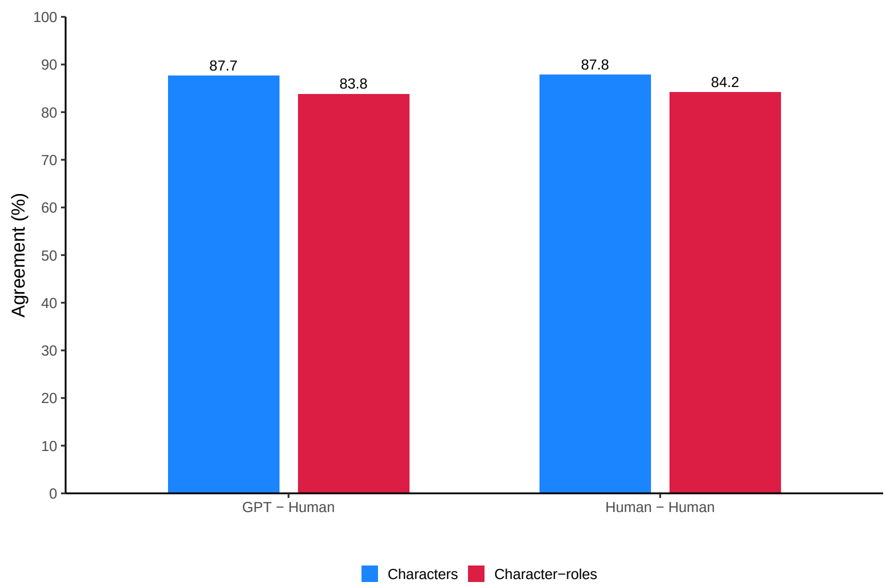
Overall, we find a high level of agreement between Human Coders and GPT. On average, GPT and Human Coders classified the presence of characters the same way in 87.7% of cases. Agreement on character-role classifications was similarly high, at 83.8%. Notably, these rates closely mirror the agreement between Human Coders themselves. Two independent Human Coders agreed on the character in 87.8% of cases and on the character role in 84.2% of cases. These results support our view of GPT as an average or representative Human Coder.

As explained above, our main agreement measure includes both positive and negative classifications. In most cases, characters are **not** present, so including negative agreement (where coders agree a character is absent) can inflate the overall agreement rates. To ensure the results are not driven by these cases, we compute Fleiss' κ, which adjusts for agreement that may occur by chance. The Fleiss' κ amounts to 0.578 for agreement on character, and 0.486 for the agreement on character-roles. These correspond to a moderate level of agreement according to the conventional interpretation by Landis and Koch (1977). Even after correcting for chance agreement and the prevalence of negative classifications, the level of agreement remains relatively high, further supporting the reliability of both human and GPT-based coding.

In the second step of our analysis, we examine the agreement on the assignment of drama triangle roles (Karpman 1968). As we argue in the paper, the drama triangle is not just a communication tool deeply rooted in the history of storytelling, but also a natural way humans interpret reality. Based on this, we expect high levels of agreement when it comes to assigning roles. In the previous analysis (Figure B.1), agreement captured both uncertainty about the presence of characters and the assignment of roles. In other words, coders were compared on both whether a character was present and which role, if any, was assigned. In the next step, we select only those tweets where the two Human Coders agreed that a specific character was present, to then compare the agreement on the assignment of roles, both between the Human Coders and between GPT and Human Coders.

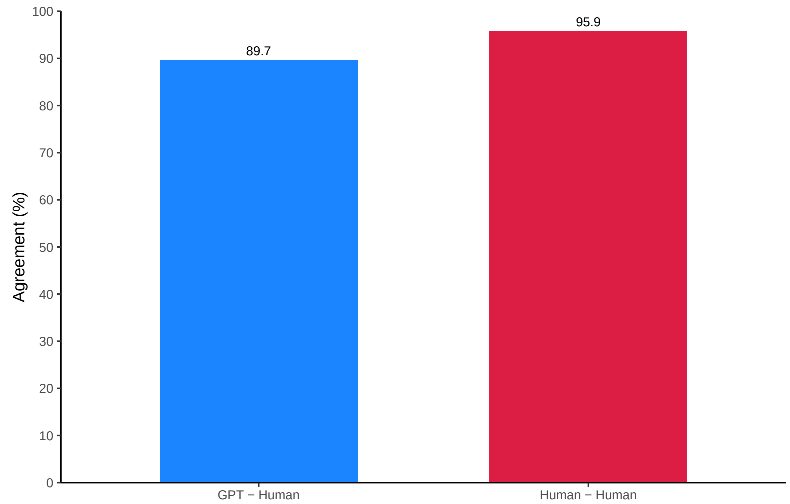
Figure B.2 shows the results of this second exercise. On the left, the blue bar indicates agreement between GPT and Human Coders; on the right, the red bar shows agreement between pairs of Human Coders, limited to tweets where both Human Coders agreed on the presence of the character. As expected, agreement levels are high: Human Coders agreed in nearly 96% of cases, while GPT agreed with Human Coders in almost 90% of cases. We compute also in this case the Fleiss' κ which measures 0.668 in this case, indicating very high agreement, as expected.

Figure B.1: ***Overall Agreement on Characters and Character-Roles***



Notes: The figure shows agreement rates between GPT and Human Coders, and between pairs of Human Coders, for the classification of 500 randomly selected tweets. All tweets were classified by GPT as containing at least one of characters of interest in our study. Twenty-eight Human Coders, each coding a different number of tweets, classified the sample. Each tweet was coded by two Human Coders. The first two bars (left) show the share of tweets where GPT and the Human Coder agreed on the presence of the character (blue) and the character-role (red). GPT is compared to whichever Human Coder classified each tweet. The next two bars show the same agreement rates between the two Human Coders who coded each tweet. Agreement includes both positive and negative cases, meaning instances where coders (or GPT and a Human Coder) agreed that a character or role was absent.

Figure B.2: ***Agreement on Character-Role Conditional on Human Coders Agreeing on the Presence of Characters***



Notes: The figure shows agreement rates between GPT and Human Coders, and between pairs of Human Coders, for the classification of 500 randomly selected tweets. All tweets were classified by GPT as containing at least one of characters of interest in our study. Twenty-eight Human Coders, each coding a different number of tweets, classified the sample. Each tweet was coded by two Human Coders. For each tweet the dataset comprises ten entries, one for each character of interest. For this exercise we retain those entries of the dataset where the two Human Coders agreed on the characters' presence. The left column shows the share of these tweets for which GPT and the Human Coders agreed on the assignment of the roles to characters. The right column shows the same for the agreement between Human Coders.

Comparison at the Tweet Level

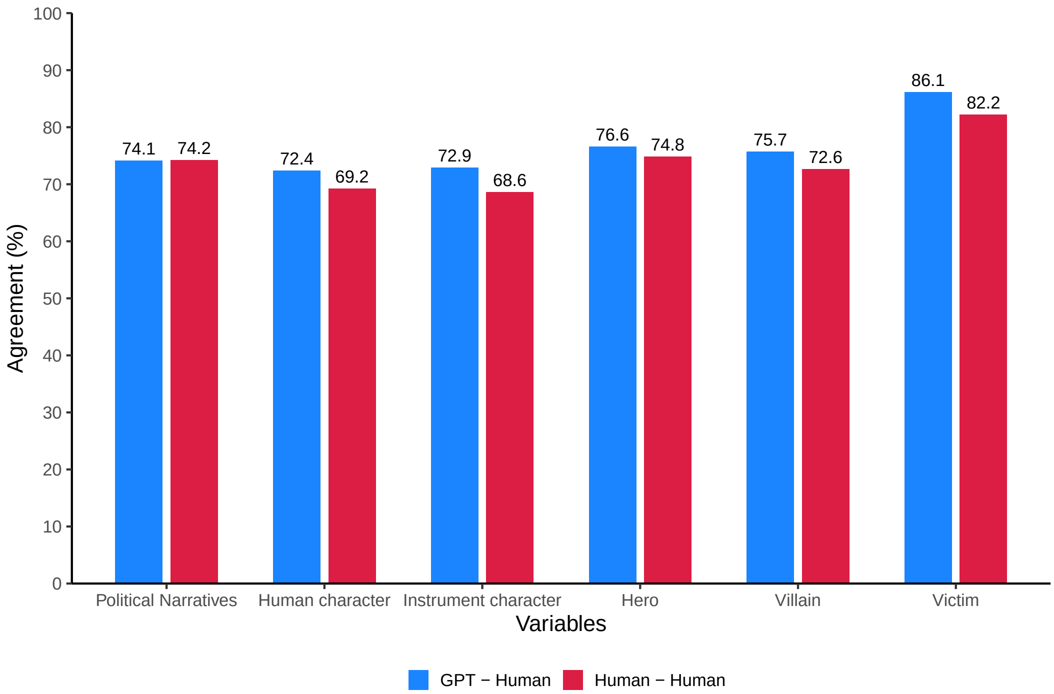
In the second part of this analysis we we compare classifications at the tweet level. This entails using the tweets classified by GPT and Human Coders, to build variables mirroring those used in the analysis, and then assess the level of agreement on these variables. Figure B.3 displays the results, on which we provide further details below.

Political Narratives We compute the *political narratives* variable at the tweet level, defined as containing at least one character-role combination. We assess agreement rates between GPT and Human Coders (blue) and between pairs of Human Coders (red). As shown in Figure B.3, agreement levels are very similar: around 74% for GPT–Human Coder comparisons and 75% for Human Coder pairs. As above, we also compute Fleiss' κ to provide an unbiased measure. The κ value is 0.215, indicating fair agreement.

Human and Instrument Characters We use the classified tweets to compute two additional variables also used in our analysis: the Human Character variable, defined as containing at least one human character, and the Instrument Character variable, defined as containing at least one instrument character. We observe a high level of agreement in detecting both human and instrument characters. As shown in Figure B.3, on average, GPT and Human Coders agreed on the presence of at least one human character in approximately 73% of tweets. Similarly, the inter-human agreement on detecting both human and instrument characters is about 69%. Notably, two Human Coders agree, on average, to a lower degree than when pairing GPT with any Human Coder. Fleiss' κ in this case is higher, at 0.43, indicating moderate agreement.

Hero, Villain, and Victim Roles Finally, we explore agreement on variables capturing the classification of roles. We construct three variables: Hero, Villain, and Victim, defined respectively as containing at least one hero, villain, or victim character-role in the tweet. This analysis further supports the validity of our approach. Figure B.3 shows that GPT and Human Coders agreed on the presence of heroes in roughly 76% of cases, villains in 76%, and victims in about 86%. Once again, Human Coders agreed with GPT more often than they agreed with each other. Fleiss' κ values are 0.508 for heroes, 0.494 for villains, and 0.266 for victims. The lower κ for victims likely reflects the lower frequency of victim roles and the uneven distribution of this classification, which often takes a value of zero. This makes it more likely to agree 'by chance' on this particular role.

Figure B.3: ***Agreement on Measures Mirroring the Variables Used in the Analysis***



Notes: The figure shows agreement rates between GPT and Human Coders, and between pairs of Human Coders, for the classification of 500 randomly selected tweets. All tweets were classified by GPT as containing at least one of characters of interest in our study. Twenty-eight Human Coders, each coding a different number of tweets, classified the sample. Each tweet was coded by two Human Coders. Agreement is defined as the number of identical classifications over the number of total tweets. *Political Narratives* is defined as the level of agreement on the presence of at least one character-role in a tweet (Hero, Villain, Victim). *Human character* is defined as agreement on the presence of at least one human character. *Instrument character* is defined as agreement on the presence of at least one instrument character. *Hero*, *Villain*, *Victim* measures agreement for the presence of at least one character-role in the tweet. The blue bars indicate the average level of agreement between GPT and the two humans. The red bar indicates the average level of agreement between two humans.

C Additional Output: Twitter sample

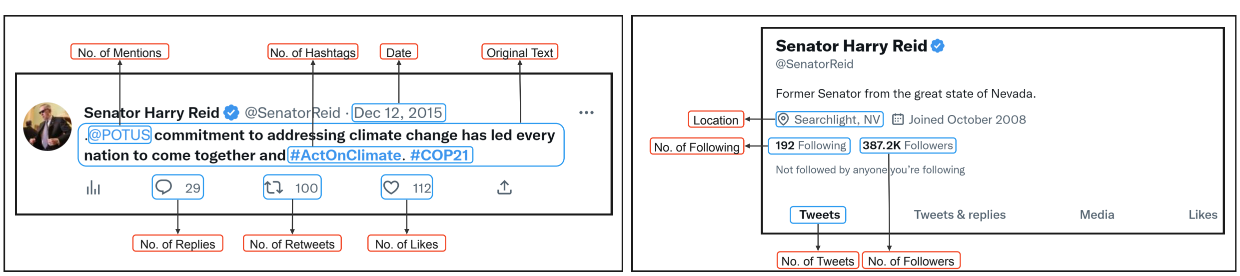
C.1 Additional Details on Variables

In this section of Appendix C, we provide additional information on the variables used in the variables used in our analysis. In particular, Figure C.1 offers a visual reference identifying the information retrieved via the Twitter APIv2 and used to construct our variables. The information retrieved is framed in blue frames, while the variables created by using that content are indicated in red frames. The left panel shows the information extracted to create the tweet related variables; the right panel shows the information extracted to create the user-related variables. Importantly, the Twitter APIv2 is no longer available for free to researchers, although some paid solutions remain accessible.

Table C.1 and C.2 list all the variables used in our analysis. For each variable, we provide a short description, the values of its categories, scale or interval, and the data source from which it was created. More specifically, Table C.1 provides information about the outcome variables, treatment variables, and the variables capturing users' features. Some important points are noteworthy about the latter. These variables were the result of own coding, done via the OpenAI API. The input for this coding exercise is the description of the profiles provided by some of the users. It is important to mention that not all users provide a profile description. Thus, for variables like religiosity, we treat equally those who had a description but did not mention being religious and those who did not provide a description at all.

Table C.2 presents the variables capturing language metrics, valence, and emotions of text, which are used in some descriptive outputs in the paper. It also includes information on date and location. The language metrics were computed directly from the text and are used to assess differences in the features of narratives and neutral tweets. These include measures such as lexical density, complexity, and vocabulary richness. Similarly, we compute measures of valence and emotions using word lists from the NRC Dictionary Mohammad and Turney (2013). These variables allow us to examine how the emotional tone and language characteristics vary across different types of tweets.

Figure C.1: *Information Scraped via Twitter APIv2*



Notes: The figure shows two screenshots. On the left a tweet posted by the user *@SenatorReid*, on the right the same user's Twitter profile. We frame all the information retrieved via the Twitter APIv2 in blue and indicate the variable for which the information is used in red frames. In Section 3 of the paper we describe our data.

Table C.1: *Description of Variables (Part I)*

Variable	Question/Description	Categories/Scale/Interval	Source
Outcomes			
No. of Retweets	Number of times the tweet is retweeted	$n \in [0; 29{,}526]$	Twitter APIv2
No. of Replies	Number of times the tweet is replied to	$n \in [0; 30{,}887]$	Twitter APIv2
No. of Likes	Number of times the tweet is liked	$n \in [0; 489{,}375]$	Twitter APIv2
Treatment			
Character-Role (predicted)	Detects whether the tweet contains a character-role presenting a character in a specific role	0 = not present, 1 = present	Own computation
Villain	Indicates at least one villain narrative is present in the tweet	0 = none, 1 = at least one	Own computation
Hero	Indicates at least one hero narrative is present in the tweet	0 = none, 1 = at least one	Own computation
Victim	Indicates at least one victim narrative is present in the tweet	0 = none, 1 = at least one	Own computation
Neutral	Indicates at least one narrative with neutral character representation in the tweet	0 = none, 1 = at least one	Own computation
Human	Indicates at least one human character presented in a role	0 = none, 1 = at least one	Own computation
Instrument	Indicates at least one instrument character presented in a role	0 = none, 1 = at least one	Own computation
Control Variables			
No. of Words	Number of words in the tweet (excluding mentions/hashtags)	$n \in [0; 117]$	Own computation
No. of Hashtags	Number of hashtags (#) in the tweet	$n \in [0; 26]$	Own computation
No. of Mentions	Number of mentions (@) in the tweet	$n \in [0; 51]$	Own computation
No. of Followers	Number of users following the tweet's author	$n \in [0; 133{,}245{,}480]$	Twitter APIv2
No. of Following	Number of accounts the author follows	$n \in [0; 4{,}066{,}970]$	Twitter APIv2
No. of Tweets	Total tweets produced by the author up to posting	$n \in [0; 9{,}611{,}963]$	Twitter APIv2
Author Characteristics			
Democrat	Indicates if the author's profile description identifies them as a Democrat	0 = no, 1 = yes	Own computation
Republican	Indicates if the author's profile description identifies them as a Republican	0 = no, 1 = yes	Own computation
Religious	Indicates if the author's profile description mentions a religious affiliation	0 = no, 1 = yes	Own computation
High Education	Indicates if the author's profile description mentions high educational attainment	0 = no, 1 = yes	Own computation
Children	Indicates if the author's profile description states that they have children	Count Variable for number of children	Own computation

Notes: The table contains a description of all the variables for outcomes (virality), treatment (character roles), control variables, and author characteristics. All data are sourced either from own computation or the Twitter APIv2. In Section 3 of the paper, we describe our data in more detail.

Table C.2: ***Description of Variables (Part II)***

Variable	Question/Description	Categories/Scale/Interval	Source
Quality of Text			
Lexical Density	Ratio of content words (nouns, verbs, adjectives, adverbs) to total words	$\in [0.01; 0.92]$	Own computation
Type/Token Ratio	Ratio of unique words to total words in the tweet	$\in [0; 1]$	Own computation
Reading Ease	Flesch Reading Ease measure (higher = easier to read)	$\in [1; 114.63]$	Own computation
Education needed to comprehend text	Approx. US grade level needed to understand the tweet (e.g., Flesch-Kincaid)	$\in [1.1; 54.27]$	Own computation
Emotions			
Joy	Joy is the average occurrence of words from the NRC dictionary associated with "joy" based on (Mohammad and Turney 2013).	Count variable	Own computation
Surprise	Surprise is the average occurrence of words from the NRC dictionary associated with "surprise" based on (Mohammad and Turney 2013).	Count variable	Own computation
Fear	Fear is the average occurrence of words from the NRC dictionary associated with "fear" based on (Mohammad and Turney 2013).	Count variable	Own computation
Sadness	Sadness is the average occurrence of words from the NRC dictionary associated with "sadness" based on (Mohammad and Turney 2013).	Count variable	Own computation
Anger	Anger is the average occurrence of words from the NRC dictionary associated with "Anger" based on (Mohammad and Turney 2013).	Count variable	Own computation
Other			
Date	Date of tweet creation	02.01.2010 – 25.12.2021	Twitter APIv2
Location	Highest level of precision at which the tweet could be located	{country, state (US), city}	Own computation

Notes: The table contains a description of all the variables related to the text metrics, valence, and emotions in text. In Section 3 of the paper we describe our data.

C.2 Additional Descriptive Statistics

In this section of Appendix C, we provide descriptive statistics on the observational data used to create the outputs in the paper and appendices. We begin by complementing the descriptive statistics shown in the paper with Table C.3. The tables reports descriptive statistics for the set of relevant tweets, those used in the analysis. For each variable, we report the mean, median, standard deviation, minimum, and maximum value, covering the level of localization, public metrics from the user's profile, and information from the profile description.

Table C.3: ***Features of Relevant Tweets (United States, 2010-2021)***

	Mean	Median	St. Dev.	Min.	Max.
Virality					
No. of Retweets (Virality)	3.7	0	154	0	29,526
No. of Likes	19	0	1,239	0	489,375
Tweet's Characteristics					
No. of Words	29	26	13	1	64
No. of Hashtags	.47	0	1.1	0	26
No. of Mentions	1.7	1	4.2	0	51
Quality of Text					
Lexical Density	.81	.82	.081	0	1
Type/Token Ratio	.93	.94	.062	0	1
Reading Ease	56	59	22	-1148	117.67
Educa. Needed to Comprehend Text	12	11	4.8	1	54.23
Emotions					
Avg. Count of Joy Words	.048	0	.081	0	1
Avg. Count of Surprise Words	.027	0	.067	0	1
Avg. Count of Trust Words	.11	.048	.15	0	1
Avg. Count of Anger Words	.048	0	.084	0	1
Avg. Count of Disgust Words	.03	0	.07	0	1
Avg. Count of Fear Words	.15	.1	.21	0	1
Avg. Count of Sadness Words	.053	0	.093	0	1
No. of Observations	309,744				

Notes: The table displays descriptive statistics for the dataset of relevant tweets used in our analysis. We define a tweet as relevant if it features at least one character from our list. We include only character roles that appear at least 100 times, thus excluding 'US Republicans-Victim', 'Emission Pricing-Victim', 'Regulations-Victim', and 'Green Tech-Victim'. For each variable, we report the average, median, standard deviation, and minimum/maximum values. We calculate the number of words per tweet excluding hashtags and mentions. We group variables by their role in the analysis. Paper Section 3 is the reference section, where we describe and discuss the data used in the analysis.

We move on with Table C.4, showing descriptive statistics about users that posted the relevant tweets. The localization level is a dichotomous variable equal to one if the user could be located, through our geo-localization pipeline, at the state level. Roughly 93% of users could be located at least at the state level. This is the most commonly used subset in our analysis, as we implement state fixed effects in most specifications and exclude users who could only be localized generically in the United States. A total of 3.4% of users' profiles were verified, at a time when verification was provided by Twitter for public figures. Users in our dataset are generally prolific: the median user has posted around 8,800 tweets (this refers to total activity since account creation, not within our dataset). In 14% of cases, users mention in their profile description that they are Democrats, and in 3.4% of cases, Republicans. Around 5% of users described themselves as religious. One quarter of users reported having higher education, and 11% reported having children.

Table C.4: ***Characteristics of Users That Posted Relevant Tweets (United States, 2010-2021)***

	Mean	**Median**	**St. Dev.**	**Min.**	**Max.**
Localization Level					
Share State-Located	.93	1	.25	0	1
Profile's Characteristics					
Share verified	.034	0	.18	0	1
No. of Followers	8,804	434	425,031	0	133,243,353
No. of Following	1,701	668	6,381	0	841,864
No. of Tweets	25,173	8,094	58,077	1	3,671,808
Profile Description					
Share of Democrats	.14	0	.35	0	1
Share of Republicans	.034	0	.18	0	1
Share religious	.056	0	.23	0	1
Share with High Educ.	.25	0	.43	0	1
Share with Children	.11	0	.31	0	1
No. of Observations	152,560				

Notes: The table provides insights into the users' characteristics for those users that posted the relevant tweets used in our analysis. We define a tweet as relevant if it features at least one character from our list. We include only character roles that appear at least 100 times, thus excluding 'US REPUBLICANS-Victim', 'EMISSION PRICING-Victim', 'REGULATIONS-Victim', and 'GREEN TECH-Victim'. For each variable, we report the average, median, standard deviation, and minimum/maximum values. We calculate the number of words per tweet excluding hashtags and mentions. We group variables by their role in the analysis. Paper Section 3 is the reference section, where we describe and discuss the data used in the analysis. Paper Section 3 is the reference section, where we describe and discuss the data used in the analysis.

Table C.4 provides insights into the characteristics of relevant tweets, defined as those containing at least one of the characters of interest in this analysis. For comparison, Table C.5 reports descriptive statistics for all tweets classified through our GPT pipeline. This broader dataset includes the relevant tweets used in our analysis, tweets classified as addressing climate change policy without mentioning any of our characters, tweets discussing the existence of man-made climate change more generally, and tweets not related to climate change at all.

Comparing the paper Table C.3 and Table C.5 some noteworthy points emerge. The subset of relevant tweets, compared to the totality of tweets, is generally more viral with retweets and like being almost twice as high. While relevant tweets are generally longer, the amount of hashtags and mentions used is comparable. Virtually all measures of text quality, emotions, and valence in text are comparable between the two datasets. Overall, the relevant tweets tend to be longer and more viral.

Table C.5: *Features of All Tweets (United States, 2010-2021)*

	Mean	Median	St. Dev.	Min.	Max.
Virality					
No. of Retweets (Virality)	2.3	0	269	0	194,217
No. of Likes	11	0	1,229	0	896,759
Tweet's Characteristics					
No. of Words	23	20	13	0	65
No. of Hashtags	.43	0	1.1	0	28
No. of Mentions	1.6	1	4.9	0	51
Quality of Text					
Lexical Density	.79	.81	.098	0	1
Type/Token Ratio	.94	.95	.062	0	1
Reading Ease	60	64	24	-2840	120.21
Educa. Needed to Comprehend Text	10	9.6	5.1	0	280.4
Emotions					
Avg. Count of Joy Words	.04	0	.082	0	1
Avg. Count of Surprise Words	.026	0	.077	0	1
Avg. Count of Trust Words	.096	0	.16	0	1
Avg. Count of Anger Words	.048	0	.1	0	1
Avg. Count of Disgust Words	.031	0	.076	0	1
Avg. Count of Fear Words	.17	.067	.25	0	1
Avg. Count of Sadness Words	.045	0	.09	0	1
No. of Observations	1,151,671				

Notes: The table reports descriptive statistics for the dataset of all tweets used in the GPT classification. For each variable, we report the average, median, standard deviation, and minimum/maximum values. We include only character roles that appear at least 100 times, thus excluding 'US Republicans-Victim', 'Emission Pricing-Victim', 'Regulations-Victim', and 'Green Tech-Victim'. We calculate the number of words per tweet excluding hashtags and mentions. We group variables by their role in the analysis. Paper Section 3 is the reference section, where we describe and discuss the data used in the analysis.

C.3 Distribution of Narratives and Virality Outcomes

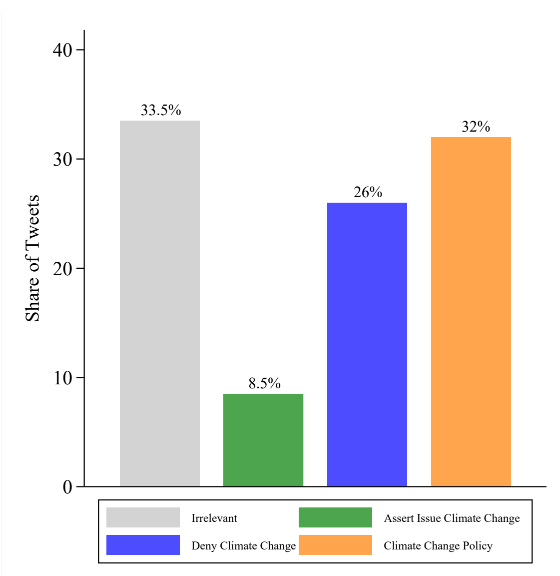
In this subsection of Appendix C, we provide additional output describing the distribution of *Political Narratives* and the virality outcomes used in our analysis. As a first step, we show how tweets were classified by GPT into the different categories defined in the two steps of our pipeline, in Figure C.2. As explained in Appendix A, in the first part of the pipeline GPT classifies tweets into four categories: *Irrelevant* (not about climate change), *Assert* (stating that man-made climate change exists), *Deny* (denying man-made climate change exists), and *Policy* (tweets about climate change policy).

Figure C.2a shows the distribution of tweets classified in the first step of our pipeline. Roughly 33% of tweets are irrelevant, meaning they are either too unclear to classify, too short, or unrelated to climate change despite mentioning keywords from our list. Tweets classified as discussing climate change policy make up about 32% of all tweets. A large share of the conversation on Twitter/X revolves around simply asserting or denying the existence of man-made climate change. These tweets do not contribute to the policy debate but instead reflect fixed and polarized positions on either side of the broader discussion.

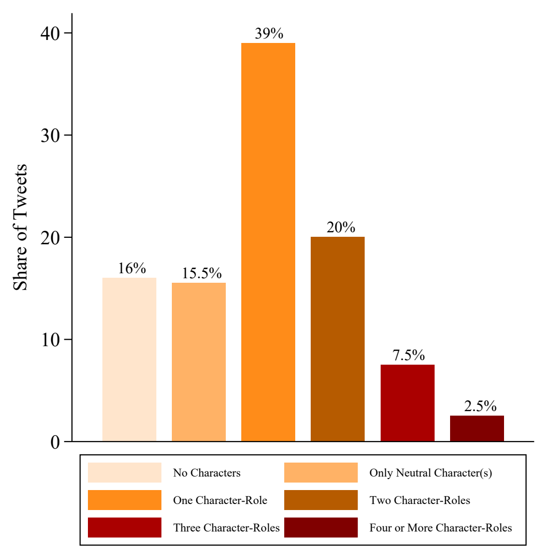
Only for tweets in the *Climate Change Policy* category do we prompt GPT to identify characters of interest. The distribution of these classifications is captured in Figure C.2b, where a clear picture emerges. The list of characters used in this analysis – developed through extensive reasoning and review of the literature and public debate – captures the discussion on climate change policy well. Our characters appear in the vast majority of tweets, with only 16% not including any of them. In 15.5% of tweets, characters are present but only in neutral form, meaning they are not assigned one of the three roles from the drama triangle. This subset serves as the comparison group in most of our analysis and forms the basis for estimating the effect of *Political Narratives*. About 39% of tweets contain at least one character-role, framing a character as a hero, villain, or victim. Two character-roles appear together in 20% of tweets, three in 7.5%, and four or more in 2.5%. Overall, the characters of interest appear in most policy-relevant tweets, with single character-roles being the most common.

Figure C.2: *Total Share of Tweets in GPT's Classification*

(a) Distribution of Policy Related and Other Tweets



(b) Total Share of Narratives by Presence of Characters



Notes: The figures provide insights into the classification obtained through our pipeline. On the left, Figure C.2a reports results from the first step of the classification, showing the share of tweets originating from the US that were labeled as irrelevant to the climate change discussion (gray), simply asserting the importance of the matter (green), denying the matter (blue), or focusing on climate change narratives about policies and solutions. On the right, Figure C.2b focuses only on the latter group of policy-related tweets. For these tweets, we show the distribution by character framing: the share containing no characters, the share featuring characters only in a neutral way, the share featuring one character-role, two character-roles, three, and four or more.

In the second part of this section, we provide additional details into the distribution of our character-roles. In other words, we dive into the classification of characters, divided into roles and neutral framing. We aim to complement the paper Table 2, which displays the distribution in shares, excluding those character-roles that did not reach at least 100 occurrences in the full set of classified tweets. We report the distributions in Table C.6.

Analyzing Table C.6, several key points stand out. As mentioned in the paper and above, some character-roles did not reach 100 occurrences: US Republicans–Victim (95), Emission Pricing–Victim (11), Regulations–Victim (29), and Green Tech–Victim (88). Although the 100-occurrence threshold is discretionary, we consider it a reasonable cutoff. We exclude character-roles below this threshold from the analysis by dropping their columns from the dataset.

Corporations and US People are the most common characters, each appearing in roughly 100,000 instances. However, many of these appear in neutral framing. The two most frequent character-roles are Fossil Industry–Villain and Green Tech–Hero, which dominate much of the public discourse. Overall, victim narratives are the least common. Only Developing Economies and US People are often framed as victims, with 5,308 and 18,113 occurrences, respectively.

Table C.6: ***Frequency of Character-Roles in Relevant Tweets (United States, 2010-2021)***

Panel A: Human Characters

	Hero	Villain	Victim	Neutral	Total
Developing Economies	777	7,025	5,308	1,166	14,276
US Democrats	33,450	10,648	333	7,924	52,355
US Republicans	683	55,568	95	9,261	65,607
Corporations	5,767	47,770	375	46,180	100,092
US People	23,450	3,221	18,113	56,833	101,617

Panel B: Instrument Characters

	Hero	Villain	Victim	Neutral	Total
Emission Pricing	11,955	7,337	11	6,985	26,288
Regulations	20,420	7,968	29	18,656	47,073
Fossil Industry	366	69,009	238	9,423	79,036
Green Tech	67,621	4,911	88	18,737	91,357
Nuclear Tech	5,077	1,879	107	2,577	9,640

Notes: The table displays the absolute frequencies of character-roles in the classified data. Sums are computed using the dataset of relevant tweets, used in our analysis. We define a tweet as relevant if it features at least one character from our list. Generally, in the analysis we perform in the paper, we exclude the character-roles that do not appear at least 100 times, thus excluding US Republicans-Victim, Emission Pricing-Victim, Regulations-Victim, and Green Tech-Victim, nevertheless we report their frequency in the table. Panel (a) displays the sums for characters of the human type, while Panel (b) displays the same for characters of the instrument type. The column Neutral in both panels reports cases where the character is present in the tweet but is not depicted in one of the three specific roles. The occurrence of character-roles is not mutually exclusive, meaning multiple roles may appear in the same tweet. The main paper Table 2 displays similar information provide insights into shares, computed excluding the categories that do not reach 100 instances.

D Heterogeneous Effects: Observational Data

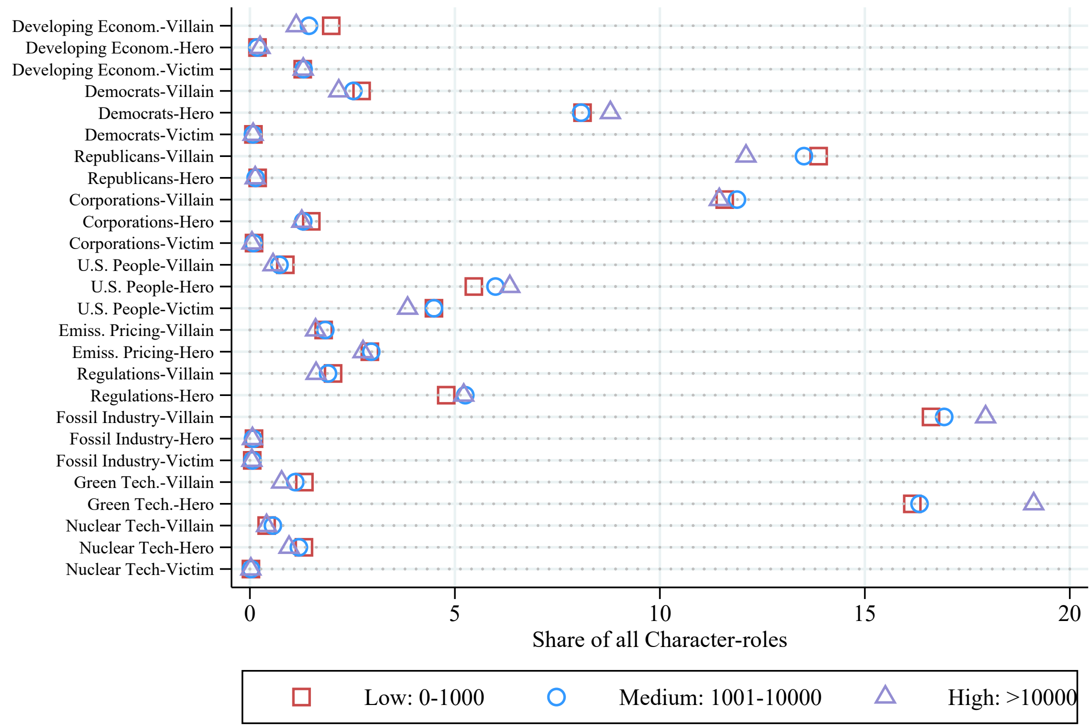
D.1 Heterogeneity by Fame of the Profile

In this paper we explore the virality of narratives in social media. We investigate what text characteristics determine the virality of tweets, through the lenses of the drama triangle. In the main part of the paper, in our analysis, we always include in our models the features of users posting tweets, such as the amount of their followers and whether they are verified users or not. In this section of Appendix D we approach the issue of virality, from the side of the users, and we ask: How does the users' fame within the platform – expressed as the number of followers – impact the kind of narratives the users spread and the virality of these narratives?

We measure the fame of a profile by the number of her followers. We divide users into three groups: low fame (0 to 1,000 followers), medium fame (1,001 to 10,000 followers), and high fame (more than 10,000 followers). As a first step, we ask whether the types of narratives differ across users with different fame levels. This check ensures that narrative content is comparable across fame categories. Figure D.1 plots the share of each character-role relative to all character-roles within each fame group. Squares represent low fame profiles, circles medium fame, and triangles high fame. We plot the shares for all character-roles that appear at least 100 times in the full dataset.

Analyzing Figure D.1, it appears clear that there are virtually no differences in narrative content across profile categories, with only a few exceptions. Low and medium fame profiles show extremely similar patterns. High fame profiles differ in four cases: they post a slightly higher share of US Democrats–Hero narratives, a considerably lower share of US Republicans–Villain narratives, a higher share of Fossil Industry–Villain narratives, and a considerably higher share of Green Tech–Hero narratives. The latter is also the most widespread narrative used by high fame users. Overall, differences across categories are rare, except that high fame users appear slightly less politicized and more concerned with energy sources. A question remains: Does this lack of difference also translate in a similar impact of featuring narratives on the virality of their tweets?

Figure D.1: ***Share of Character-Roles by fame of the Profile***

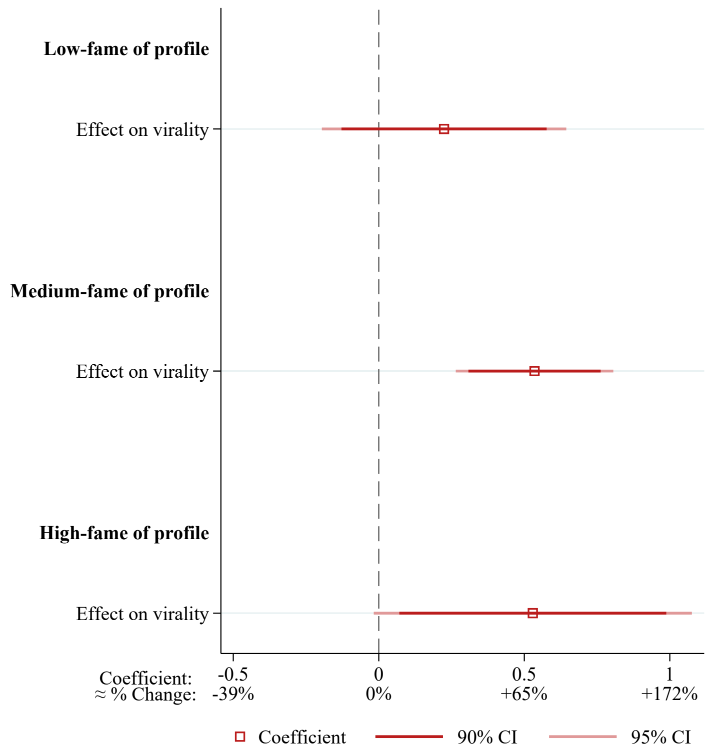


Notes: The figure shows the share of character-roles featured in relevant tweets by profile fame. Squares represent profiles with fewer than 1,000 followers, circles represent those with 1,001–10,000 followers, and triangles represent profiles with more than 10,000 followers. Shares are computed by dividing the number of times each character-role appears in tweets from a given profile category by the total number of character-roles used within that category.

After showing that narrative content is similar across profiles with different fame levels, we turn to the next question: Does this similarity lead to a similar impact of narratives on the virality of tweets across these groups? In other words, building on the main empirical results of the paper – that show narratives increasing the virality of tweets compared to neutral framing of the same characters – we now ask whether this effect holds within different types of profiles. Figure D.2 provides insights on this question. The coefficients plot shows the impact of containing a narrative on virality, the count of retweets, compared to featuring characters in neutral framing. The results are based on a slightly modified version of the main specification used in the paper. We use a Poisson Pseudo-Maximum Likelihood regression model, including character fixed effects, and hour, week, and year-state fixed effects. We do not include author characteristics, to avoid capturing variation that defines the fame categories themselves.

This exercise provides some interesting results. Despite larger confidence intervals for the high fame group – due to a smaller number of users – medium and high fame profiles show similar patterns. In both groups, narratives increase virality in line with the main results of the paper. In contrast, tweets from low fame profiles show no measurable impact of narratives on virality. This finding lends itself to several interpretations. The most likely is that at low levels of fame, the content posted by these users rarely goes viral, regardless of whether it contains a narrative or presents neutral framing. In other words, at low follower counts, factors other than content – such as limited reach or engagement – may prevent virality, making the narrative effect negligible.

Figure D.2: ***Regression Results - Impact of Political Narratives on Virality by fame of Profile***



Notes: The figure shows the coefficients of Poisson Pseudo-Maximum Likelihood regression models testing the effect of featuring at least one character-role vs. featuring characters only in a neutral role on virality, measured as the count of retweets. We explore the effects in three sub-populations: top panel, for profiles with fewer than 1,000 followers, middle panel for users those with 1,001–10,000 followers, and bottom panel for profiles with more than 10,000 followers. The x-axis reports coefficient estimates, 90% confidence intervals (dark red), and 95% confidence intervals (light red). We label also the corresponding percentage change rounded to the closest unit and computed as follows: $\approx e^{\beta} - 1$. All regressions control for author characteristics (verified status, number of followers/followings, total tweets created, party affiliation, religiosity, higher education, and parenthood) and include character fixed effects. We also include hour, week-of-year, and year–state fixed effects. Standard errors are clustered at the week level, covering the full time frame.

A note of caution is necessary for this analysis. The information on followers count reflects the amount of followers at the moment of data extraction, which means it is an ex-post measure relative to when the tweets were posted. It is therefore possible that the fame of some profiles results from the use of narratives, rather than causing it. While this is the best available approach given the data, the results should be interpreted carefully.