Communicating Europe: a computational analysis of the evolution of the European Commission's communication on Twitter

Roberta Rocca[1] · Katharina Lawall[2] · Manos Tsakiris[2,3] · Laura Cram[4]

Abstract

Social media is an important means of communication for political agencies, which makes it possible to engage with large sectors of the public. For institutions which are not directly elected by voters, such as the European Commission (EC), social media can be a strategic tool for increasing perceived legitimacy and citizen engagement, especially in contexts of high politicization. In this paper, we use natural language processing techniques to provide a comprehensive overview of how EC communication on Twitter has evolved between 2010 and 2022, with respect to both its topics and its style. Our analyses show that, over time, the focus of EC communication has shifted substantially from economy-, finance- and governance-related topics, towards social policy, digital and environmental policy, and identity. These changes have progressively differentiated the EC's profile from that of other institutions (especially more technocratic ones) and contributed to better alignment with engagement patterns of its social media audience. In addition, EC communication has become less neutral (in favor of more positive sentiment), simpler, and more readable, all features which are associated with more accessible and engaging messaging. Yet, while the EC currently scores better than most other reference agencies on several descriptors of accessibility, its style is still lexically more complex, less concrete and less action-oriented than that of other institutions. Alongside providing novel insights on how the EC's online communication and projected political identity have changed over time, this study lays the foundations for future experimental and hypothesis-driven work combining social media data with external data sources.

Keywords Political communication · Natural language processing · Twitter · European Commission · Politicization

Extended author information available on the last page of the article

🖄 Springer

Introduction

For executive bodies, social media is a primary means of communication and engagement with the general public. In the light of increasing politicization of European integration [1–3], social media could provide an important communication channel for European institutions, as their wide outreach in theory allows them to actively shape public perception and influence debates on European integration [4–6].

Supranational executive bodies like the European Commission (EC) can especially benefit from being able to directly reach out to the wider European citizenry through social media. The European Commission has always occupied a unique institutional space, often referred to as a missionary bureaucracy [7], tasked formally with a political role as guardian of the European Union treaties, as well as the traditional technocratic functions associated with its executive status. Social media provide strategic channels to increase everyday communication with citizens. In the absence of other direct accountability mechanisms, these channels can provide an opportunity for institutions like the EC to legitimize their activities in the public eye. They also allow the EC to contribute to, and potentially shape debates about the role that the EU plays in the lives of its citizens and perhaps even to influence emerging conceptions of European identity. Yet, a detailed analysis of how the EC has used the strategic potential offered by social media, and how its communication has evolved, is currently lacking.

Recent research has provided an important and relevant bird's eye perspective on the social media communication of EU supranational actors [5]. In this paper, we build on this work and extend it, zooming in on the temporal evolution of topics and style of social media communication of the European Commission as the EU's executive institution. We also extend Özdemir and Rauh's [5] work by showing how the thematic focus of the European Commission's communication has evolved over time.

Leveraging advances in public availability of social media data (e.g., access to historical data through the v2 Academic Twitter API) and state-of-the-art text modeling (e.g., [8–10]), we conduct a comprehensive analysis of the content and stylistic features of the European Commission's communication on Twitter (**@EU_Commission**) between June 2010, when the official account was created, and July 2022. By analyzing how topics and stylistic traits of EC Twitter communication have evolved, and benchmarking the EC against a range of other institutions, we aim to understand how the EC has reshaped its projected institutional identity over the past decade, and how/whether its communication has adapted to better resonate with the European public.

Research questions

Our study is structured as follows. First, we extracted daily tweet volumes between 2010 and 2022, and compared the EC's tweet volume with that of a range of reference agencies, including: (a) national governments with English as official language

(UK and Scotland); (b) other legislative and executive EU bodies (European Parliament and Council of the European Union); (c) transnational economic and monetary institutions (European Central Bank and International Monetary Fund); (d) other transnational intergovernmental organizations (the United Nations and the Organisation for Economic Co-operation and Development, OECD).

Secondly, we extracted prominent topics in the EC's Twitter communication and analyzed how the volume of tweets for each has changed over the years. Thirdly, we analyzed how the style of the EC's tweets has evolved, especially focusing on style features that are associated with increased accessibility and engagement [11, 12], and compared the EC's style with that of other reference agencies. Finally, we used predictive models to systematically analyze how topics and stylistic traits relate to engagement, and gain additional insights into which features of the EC's Twitter communication resonate most with its public.

Through these analyses, our study addresses the following research questions:

1. To what extent has the European Commission used Twitter as a communication channel over the past years? How does the amount of tweets produced by the EC compare to that of other supranational and national agencies?
2. What are the main themes of EC's Twitter communication? How have they evolved over time? Can this evolution reflect strategic attempts to generate higher engagement and support from its public?
3. How has the linguistic style of EC's Twitter communication evolved over time? Is there evidence of it evolving towards more engaging and accessible messaging, a prerequisite for successful communication and increased self-legitimation? How does the style of EC communication on Twitter compare to that of other agencies—more specifically, is it less accessible than that of other agencies (as found for more traditional communication channels)?
4. Which topics and features of style relate positively to engagement from EC's Twitter audience? Do expected relations between individual features of linguistic style and engagement hold for the EC public? Do predictive models lend evidence to the hypothesis that dynamic changes in topic and style of EC communication may have been beneficial to public engagement?

Theoretical significance and related work

By addressing these research questions, this paper makes a contribution to our understanding of supranational institutions' digital communication strategies in general, and the European Commission's communication on Twitter in particular. Beyond work by Ozdemir and Rauh [5], this paper contributes to a more general literature on government agencies', and in particular, supranational institutions' digital communication. Supranational institutions' public communications matter because they present an opportunity for institutions to craft narratives about who they are and what their purpose is [13].

This is particularly important for the EC in times of heightened politicization, which refers to the idea that European integration has become the subject of "increasingly salient and polarised public debate among an expanding range of actors" (p.10) [1]. In this context, digital communication presents a tool for self-legitimation, whereby the EC tries to communicate to the public that it has a legitimate right to authority [5]. Research on public diplomacy has highlighted its increasing "digitalization", which describes the long-term process through which digital technologies are influencing the "norms, values, working routines and structures of diplomatic institutions, as well as the self-narratives and metaphors diplomats employ to conceptualize their craft" [14]. Shifting communication modes and platforms therefore have implications both for the institutions' audiences and subjects, and for the institutions themselves (see also [15]).

Research on digital diplomacy has investigated how the European Union uses digital tools to enhance its reputation and legitimise its policy objectives [16]. Conducting interviews with European External Action Service (EEAS) officials, Hedling (2020) finds that storytelling is an important part of EU public diplomacy, and a response to an increased need to create greater buy-in for EU policies among domestic audiences. A key objective of digital storytelling is audience engagement, rather than just passive information dissemination. This highlights the strategic importance of audience engagement as a tool and a goal for supranational actors like the EU.

Building on this research, one of the key research questions that this paper addresses is in how far the European Commission's Twitter communication has evolved towards a more engaging linguistic style. In another recent application of the strategic narratives framework to the EU's online communication using quantitative text analysis methods, Moral [17] examines how the EU used different narratives to manage its reputation during the 2020 outbreak of COVID-19 pandemic. Similar to this work, we use topic modelling to understand what key themes emerge and evolve in the EU's online communication over time.

Finally, research on the EU's digital communication has also pointed out instances where key narratives, such as on gender equality, are less prominently featured in the EU's discourse than expected [18]. Similarly, with our over-time analysis of key themes in the European Commission's Twitter communication, we contribute to a broad understanding of which theoretically foundational themes (such as the ones relating to identity) are absent from the EC's discourse until more recent years.

Data sources and rationale

Our analyses are conducted on the full volume of tweets in English produced by the European Commission's English Twitter account (**@EU_Commission**) and by a number of reference agencies. We selected two national executives (UK government, **@10DowningStreet**; Scottish government, **@scotgov**, selected because

English-speaking), other legislative and executive EU bodies (European Parliament, **@Europarl_EN**; Council of the European Union, **@EUCouncil**), transnational economic and monetary institutions (European Central Bank, **@ecb**; International Monetary Fund, **@IMFNews**), and other transnational intergovernmental institutions (the United Nations, **@UN**, and the Organisation for Economic Co-operation and Development, **@OECD**). Our final database includes the full volume of tweets produced by the European Commission and each of these reference entities up to the beginning of the present study in July 2022.

Note that the decision to include only the European Commission's main official account, rather than including, for example, tweets from individual members of the European Commission or Directorates-General (DGs), is motivated by a number of considerations. First, we were interested in capturing discourse from the European Commission in a broad and comprehensive fashion, rather than identifying granular trends in specific policy areas. Secondly, contrary to accounts of individual representatives, the European Commission's main account provides uninterrupted data for the entire time span. Thirdly, focusing on the **EU_Commission** account ensures that views expressed in the tweets reflect views of the European Commission, rather than views of individual representatives. We acknowledge that our results might partly be a reflection of these choices, and that the patterns observed in our analyses may also be influenced by evolving social media management strategies related to the European Commission's main official account. Analyses focusing on tweets from a more distributed set of sources might yield results that could fruitfully complement our work.

As to the choice of reference agencies, the rationale for selection was the following. We selected a range of agencies which would: (a) provide a suitable volume of tweets in English; (b) be *comparable* to the EC under different respects. The two national executives were selected because they are two English-speaking governments located in the same geographical area (note that the US government does not have). The UN and the OECD were selected because they are international organizations with comparable breadth in policy areas. Conversely, the ECB and the IMF were selected as a sample of institutions with technocratic focus (the former being a European institution too). While this is far from a representative sample of all comparable institutions, it provides a rich and composite benchmark against which EC communication can be compared and interpreted.

Data were accessed through the v2 Academic Twitter API (now discontinued), and the text of tweets was preprocessed following a standardized pipeline (see section "Further methodological details"). Tweet metadata such as time of creation and engagement metrics (likes, retweets, quotes, replies) were also downloaded. Engagement data were condensed into a single engagement metric (which we will henceforth refer to as "engagement") by summing the number of likes, retweets, quotes, and replies. For all analyses including engagement metrics, we use raw engagement data, as the Twitter API does not expose the number of visualizations (which is needed to compute engagement rates) nor historical data on other potential normalization factors such as follower counts.
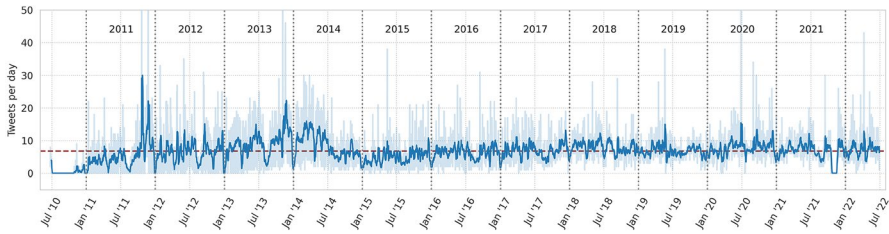
Fig. 1 Number of tweets per day produced by the EC's Twitter accounts between July 2010 and July 2022. The light blue line represents the raw count of tweets for a given day. The darker line represents temporally smoothed estimates. Smoothing is performed by computing a rolling average over 7-day windows. The red horizontal line marks the average number of tweets per day over the full time span

Methodology and results

Twitter use by the EC

To tackle our first research question on how much the EC has used Twitter as a communication channel compared to the other reference agencies, we computed the raw volume of tweets produced by each of these agencies since the time their official accounts were created. Table 1 displays the overall number of tweets and the number of tweets per day for each of the target entities (EU_Commission highlighted in bold). As the table shows, the EC ranks second among our selected entities, with 6.75 tweets per day, suggesting that the EC has made conspicuous use of this communication channel over the years.

Figure 1 displays the daily number of tweets produced by the EC between 2010 and 2022. After peaking in 2013 and 2014, the number of daily tweets (and its variability) have become fairly stable and revolving around the overall mean. More detailed visualizations of tweet volumes for reference agencies are displayed in the Appendix.

Main themes in the EC's communication

Topic model

To answer our second research question on how the main themes in EC communication on Twitter have evolved, we used contextualized topic models [8, 19] to extract a set of interpretable topics from the full corpus of EC tweets. Topic models are a data-driven computational method that can be used to extract semantic themes commonly occurring in a large set of texts based on word co-occurrence statistics. A topic is, in fact, simply a set of words that tend to co-occur in the same texts, defining "semantic attractors" in the set of texts under analysis. Compared to traditional topic models, contextualized topic models leverage pretrained transformer models [20] to improve coherence and interpretability. To estimate the topic models, we implemented a robust model selection procedure and selected a pretrained

DistilBERT model with 20 target topics and a 500-word vocabulary (see section "Further methodological details" for more information).

Our topic model identified the following core topics (displayed in Table 2), which we labeled by inspecting the top 10 topic-defining words and the 10 most representative tweets for each topic (see Table 17, provided in the Appendix), and we clustered into seven macro-categories:

- **Economic and financial policy**, including: Economy and Markets; Finance and Trade; Growth and Global Development; Strategic Investments (e.g., recovery, research, and innovation);
- **Social policy**, including Health; Citizens' rights and integration; Human rights;
- **Environmental and digital policy**, including: Digital policy; Digital and green transition; Energy, Sustainability and Climate;
- **Identity and citizen participation**, including: Identity, Culture, and Citizen Engagement; Visions for the Future; Citizen Initiatives;
- **Governance**, including: Internal governance; Trade, partnerships, and law;
- **Solidarity and humanitarian aid**, including: Solidarity and emergency response; Humanitarian aid;
- **Communications and media**, including Press conferences and official statements; Charts, links, and infographics; Live events.

How have topics evolved?

We used the topic model to quantitatively describe the thematic content of each tweet in the dataset. For each tweet and topic, we extracted a score quantifying the extent to which a given topic is represented in the tweet. Based on these scores, we classified tweets as belonging to one of the 20 topics identified by the model, by simply extracting the topic whose score is highest. Aggregating the volume of tweets for each topic over time yielded a time series that makes it possible to understand how the focus of EC communication on Twitter has evolved over the entire time span.

Figure 2 (left panel) displays the resulting aggregated data. The heatmap shows that, while up until 2016 EC Twitter communication focused heavily on economy-related topics, governance, and sharing of information on institutional events and dynamics, 2017 marks a turning point. From 2017 on, tweets on social policy, environmental and digital policy, identity and citizen engagement become prominent. Health becomes especially prominent in correspondence with acute phases of the COVID-19 pandemic (Fig. 3). At a more fine-grained level, environmental policies display notably higher tweet volume from 2020 on, while digital policy, citizens' rights and integration, and human rights (prominent in 2018 and 2020) have witnessed a slight decrease after 2020. The number of tweets related to solidarity and humanitarian aid have increased dramatically in 2022, arguably a reflection of Russia's invasion of Ukraine. The right panel of Fig. 2 zooms in on the volume of tweets per topic over the past 3 years. *Health*, *Energy*,

Table 1 Overview of the tweet volume for each institution

	# Tweets	First tweet	Tweets per day
UN	42,664	20-12-2008	8.46
EU_Commission	**29,621**	**2010-06-23**	**6.75**
IMFNews	19,361	2009-03-24	4.00
Europarl_EN	18,458	2009-04-30	3.84
ecb	14,948	2010-04-28	3.36
OECD	13,558	2009-03-19	2.80
scotgov	12,090	2009-06-16	2.54
10DowningStreet	11,595	2008-03-26	2.23
EUCouncil	5625	2012-12-04	1.62

Sustainability and Climate, *Digital Policy*, and *Visions for the Future* have been the most prominent topics on average between 2020 and 2022. On the other hand, *Solidarity and Emergency Response*, and *Identity, Culture, and Citizen Engagement* have been the most prominent topics in 2022.

Estimates of the volume of tweets over time (Fig. 3) lend further support to these patterns. Focus on economy and governance-related topics, as well as on institutional events and infographics, has decreased steadily over time. Tweets on social policy, environmental and digital policy, and identity and citizen participation have increased steadily over time. Tweets on solidarity and humanitarian aid have steeply increased over the past year. Note that these trends are fairly consistent across individual topics within each topic category.

Comparison with other agencies

The evolution in the amount of space given to different areas of policy on Twitter has radically altered the content profile of EC communication on Twitter. How do these changes affect the distance of the EC to other agencies, in a space that includes both technocratic agencies like the ECB, other European institutions, and elected national governments? To answer these questions, we quantified the similarity between topic volumes in EC Twitter communication and topic volumes for other agencies over time. We did so by computing month-by-month pairwise correlations between the proportion of tweets produced, for each topic, by the EC, and the corresponding values for other reference agencies (Fig. 4). Higher correlations denote a closer match between the two agencies in the amount of focus given to each of the 20 topics across.

The resulting values are displayed in Fig. 4. The data shows two notable trends. First, the focus of EC communication has become increasingly dissimilar to that of more technocratic institutions such as the ECB and the IMF. Secondly, although less dramatically, EC communication has also become more distinct from that of national governments in our sample. More complex temporal fluctuations, with an overall decrease in similarity, can be observed for the remaining

Table 2 List of topics and topic-defining words extracted through contextualized topic modeling

| Category	Topic name	Topic-defining words
Economic and Financial policy	Economy and Markets	Market, euro, tax, area, business, emissions, annual, financial, sector, prices
	Finance and Trade	Commission, area, euro, states, trade, member, EU, aid, rate,annual
	Growth and Global Development	Global goals, money, climate action, Asia, emerging, governments, migration, needed, income, impact
	Strategic Investments (e.g., recovery, research, innovation)	Support, million, investment, projects, recovery, billion, financing, plan, new, research
Social Policy	Health	Vaccines, coronavirus, vaccine, people, covid, pandemic, help, spread, safe, need
	Citizen Rights and Integration	EU, new, rules, rights, countries, European, citizens, information, online, safety
	Human Rights	Women, rights, violence, equality, human, online, girls, every,right, children
Environmental and Digital Policy	Digital Policy	EU, new, rules, data, online, ensure, rights, countries, protection, access
	Digital and Green Transition	Energy, new, climate, digital, Europe, sustainable, green, transition, strategy, economy
	Energy, Sustainability and Climate	Energy, climate, europe, sustainable, digital, economy, planet, green, make, emissions
Identity and Citizen Participation	Identity, Culture, and Citizen Engagement	Europe, European, today, together, future, join, day, year, us, time
	Citizen Initiatives	European, EU, initiative, europe, today, citizens, day, years, take, know
	Visions for the Future	Europe, European, future, today, climate, change, together, ideas, us, join
Governance	Internal Governance	President, EU, European, commission, statement, summit, council, euco, joint, speech
	Trade, Partnerships and Law	EU, commission, agreement, trade, countries, member, states, rules, law, new
Solidarity and Humanitarian Aid	Solidarity and Emergency Response	EU, support, million, humanitarian, aid, Ukraine, assistance, countries, help, billion
	Financial aid	Million, humanitarian, support, aid, help, assistance, emergency, refugees, funding, fund
Communication and Media	Press conferences and statements	President, live, press, conference, junckereu, cet, statement, commissioner, speech, follow
	Charts, links, infographics	Globalgoals, Asia, climate action, week, latest, governments, local, agenda, info, check
	Live events	Live, conference, press, cet, watch, junckereu, following, statement, follow, meeting
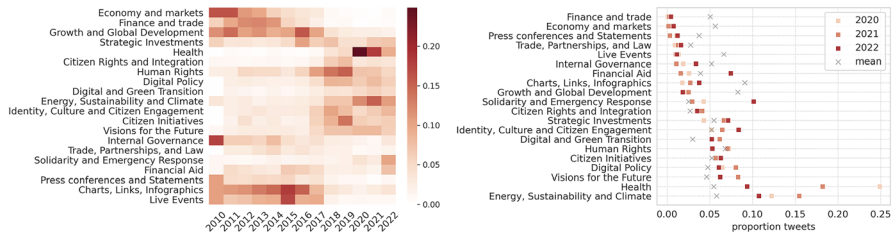
Fig. 2 Left: Topic volume for each of the 20 topics identified by our topic model, over the entire lifetime of the EC's Twitter account. Colors represent the proportion of tweets, within a given year, that score highest on the target topic. Right: proportion of tweets per topic between 2020 and 2022. Average across all years is also displayed for reference

agencies (see Appendix for a more detailed visualization). This suggests that, as a result of the changes in relative focus on different areas of policy described in the previous section (be it the result of a shift in policy priorities or of a shift in communication strategies), the content profile of EC communication has become highly distinct from technocratic institutions, and generally more distinct from other agencies, including national governments.

Alignment with audience engagement

Changes in focus could arguably reflect—at least in part—changes in policy priorities and in the ways the European Commission presents its identity. Both changes in policy priorities and in communication strategy may be a reaction to increasing politicization, and an attempt to elicit more engagement from citizens and increase self-legitimation. If this is the case, and if the strategy is at least partly successful, over time, topics of EC communication should align better with preferences of the EC's online audience.

To explore this question, we correlate the relative ranking of each topic in the volume distribution (i.e., how often each topic is tweeted about at a given time, relative to other topics) with its ranking in the engagement distribution at previous time points (i.e., how many engagements a given topic elicits on average, relative to other topics). We do so for every month between June 2010 and July 2022, to extract an estimate of how much the content produced by the EC on Twitter aligns with previous patterns of user engagement. Higher correlation values correspond to a high degree of alignment between EC contents and audience topic engagement. If correlations values increase over time, this can be interpreted as evidence of EC communication becoming increasingly more aligned with its social media audience. To test this hypothesis, we fit simple regression models with time as a regressor and correlation-based alignment estimates as outcome variable for each agency in our sample.

The results show that EC communication has indeed progressively evolved towards topics that better resonate with its social media audience ($\beta = 0.0039$, $p < 0.001$). Notably, the EC has the highest regression coefficient among all institutions in our sample, suggesting that alignment between topic volumes and engagement patterns has increased at a faster rate in the EC's communication, compared to
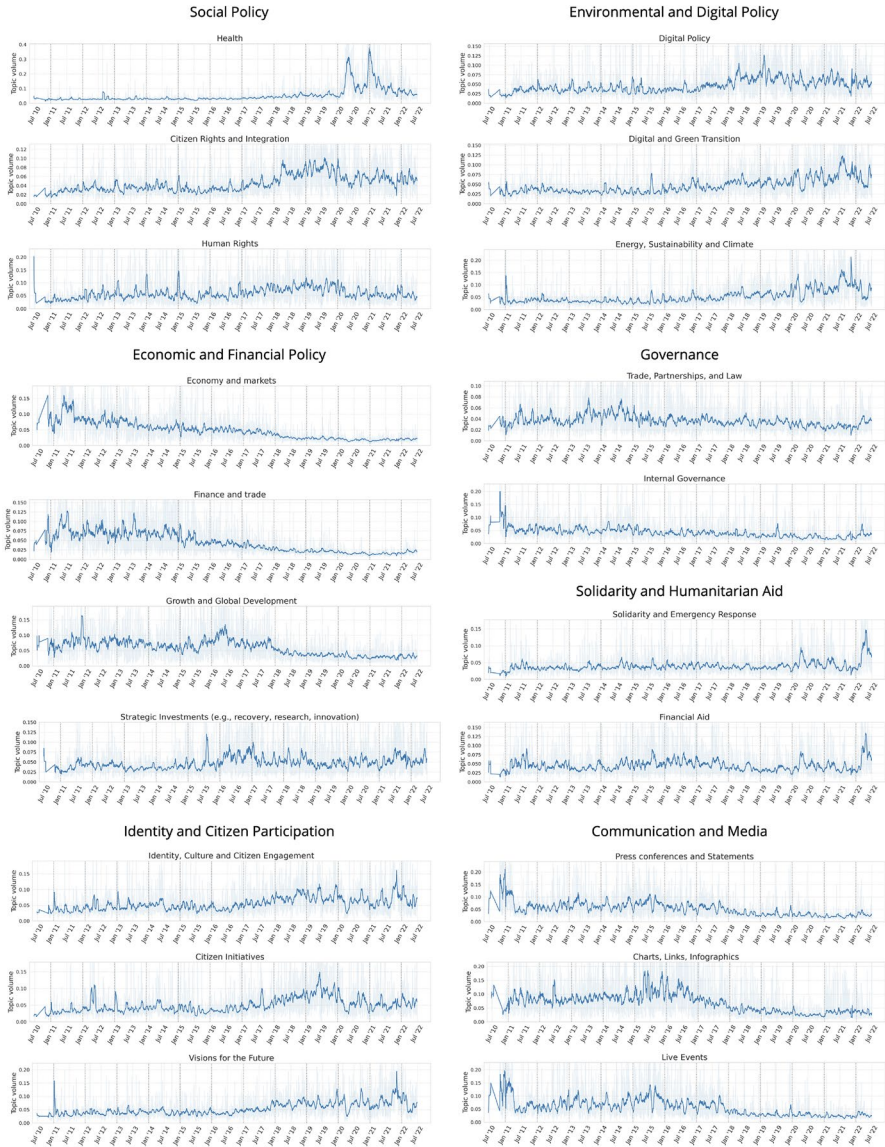
Fig. 3 Topic volume over time for each topic. Darker lines represent smoothed data (rolling average over 21 days). Lighter lines represent raw data per day

other institutions. Full reports and visualizations of these results are provided in the Appendix.

Note that engagement counts reflect preferences of the EC's *own* public, which is not fully representative of the general public. Self-selection processes play a role in both followership processes and presence on a social media platform. The

interpretation of results involving engagement counts should account for these processes. For example, in this context, it is important to highlight that alignment between EC topics and engagements is reflective of the EC's increased ability to generate content that resonates with its own audience, but it does not necessarily reflect increased ability to generate engagement in the general public.

Linguistic style of the EC's communication

Linguistic style and engaging messaging

In previous paragraphs, we analyzed EC communication focusing on its content, both in a diachronic perspective and in relation with other agencies. We showed that EC communication has drifted away from its initial focus on economy-, finance- and governance-related topics, and shifted towards areas of policy (social policy, health, digitalization, solidarity, identity) that generate more public engagement and make its profile markedly distinct from that of technocratic institutions. We also noted that 2017 seems to mark a turning point, which we speculate could be a reaction to the Brexit referendum.

These results could be compatible with the idea that the EC's Twitter communication may have evolved as a response to increasing politicization, using online communication as a means to increased self-legitimation, a process by which "authority holders engage in nurturing the belief in their claim to rule among relevant audiences" (p.134) [5]. To successfully increase self-legitimation, communication should be sufficiently engaging and accessible to the general public also from the point of view of linguistic style [5]. However, public communication from the EC on traditional media has been argued to be much less comprehensible than that of comparable institutions [12]. While we do not claim to make any inferences about the relationship between politicization and the described changes in the EC's Twitter communication in this paper, future research should test these relationships.

To address our third research question on the linguistic style of the European Commission's Twitter communication, we focus on describing how it has evolved over time and how it compares to that of other reference agencies. In doing so, we analyze two main classes of features: sentiment, a feature which has been shown to be related to engagement [21–23] and a broad feature set including several
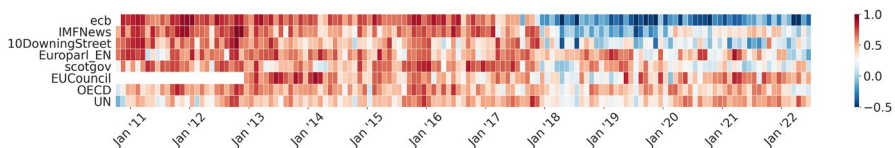
Fig. 4 Similarity between topic distributions in EC's Twitter communication and other agencies. Similarities (operationalized as rank correlations between distributions of topic volumes) are calculated on aggregate monthly topic volumes

markers of readability, complexity, and use of platform-specific communication tools like emojis, hashtags and mentions. By tracking how these features have changed over time, and benchmarking the EC against other agencies, we seek to understand whether EC communication has evolved towards more engaging and comprehensible messaging, and how its present style compares to that of other agencies.

Features

To comprehensively describe readability, complexity, and platform-specific stylistic features, we use a number of fine-grained descriptors. These include proxies for word-level or lexical complexity (the average frequency of words in corpora of standard English, see [24], as well as average word length); proxies for sentence-level complexity (the number of words or characters in a sentence); features quantifying the extent to which text is action-oriented (e.g., the verb-to-noun ratio), with the rationale that more verbs and fewer nouns yield simpler and more accessible sentences [12]; compound readability indices [25–31]; the frequency of hashtags, mentions, emojis by tracking the present of distinctive non-alphabetical markers (#, @, unicode emoji codes); average lexical concreteness [32]. To extract these metrics, we use the Python packages TextDescriptives [33], and pliers [34]. Sentiment is quantified using a pretrained sentiment classification RoBERTa model [35] fine-tuned on sentiment classification for Twitter text ([36]; available at https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment).

Sentiment

Figure 5 displays how the sentiment of EC tweets evolved over time, also compared to other agencies. The proportion of tweets with positive sentiment increased significantly from 2017 on, becoming markedly higher than the proportion of positive tweets produced by all other agencies. This is accompanied by a corresponding decrease in tweets with neutral sentiment, with the EC producing markedly fewer neutral tweets in recent years than all other agencies. The proportion of negative tweets remains roughly constant and comparable to other agencies, with the exception of a slight increase following Russia's invasion of Ukraine.

We tested these patterns statistically, by fitting a robust regression model (using Huber loss to account for the presence of outliers, [37]) with a categorical variable coding for the agency authoring the tweet as predictor, and normalized sentiment scores as outcome variables. We fitted one model for each year, to highlight how differences between agencies have changed over time. The sign of coefficients provides information on whether a given agency tends to score higher or lower on the target metric, and their values provide information on the magnitude of the effect. The results are displayed in Fig. 6. Outcome variables were standardized before model fitting, to enhance comparability across features. Results for 2022 and immediately preceding years clearly show that EC communication is on average more positive, less neutral, and slightly less negative than that of all other reference agencies.

Readability and complexity

Figure 7 displays how the linguistic style of EC communication has evolved over time, focusing on a range of features that quantify text complexity at different levels. Note that complexity, here, refers to factors influencing the amount of effort and knowledge needed to process and understand a given text, and the dimensions chosen as proxies of complexity build on methodologies from previous studies [11, 12]. We extracted: an aggregate index quantifying overall readability ("reading complexity", which synthesizes multiple readability indices using principal component analysis, see *Further methodological details* for more information); tweet length (synthesizing character, syllable, and word-level measures of overall length through principal component analysis, see section "Further methodological details" for more information); sentence length (a summary variable aggregating sentence length in characters and in words, extracted using principal component analysis, see section "Further methodological details" for more information); action-orientedness, quantified as verb-to-noun ratio [12]; indices of word-level complexity, including both the average length of words in a tweet ("word length", extracted by aggregating syllable and character count per word using principal component analysis, see section "Further methodological details" for more information) and the average frequency score of words included in the tweet (quantified as Log10 frequency in the SubtlexUS corpus, [24]—note that more frequent words involve less complexity, as they are more familiar and refer to more common and simpler concepts); average lexical concreteness [32]; the ratio of emojis, hashtags and mentions over words in a tweet.

We observe the following patterns. Overall reading complexity has decreased slightly over time. A notable decrease can be observed for word-level length, accompanied by a slight increase in average word frequency. Verb-to-noun ratio has also increased slightly. These patterns suggest an overall decrease in lexical complexity and increased action-orientedness. On the other hand, tweet-level and sentence-level complexity have increased over time (arguably a reflection of Twitter's extension of character limits in 2018), and concreteness has decreased. The use of hashtags and mentions has also decreased, while use of emojis has become more prominent.

These tendencies show that increased tweet- and sentence-level complexity is accompanied by decreased lexical complexity and increasing use of multimodal communication tools like emojis, which may result in simpler and effective communication. But how does the linguistic communication style resulting from these changes compare to that of other agencies? Is EC style overall simpler and more accessible?

To answer these questions, we performed pairwise statistical comparison between the EC and all other agencies for all the features analyzed in this section. Similar to the analysis performed for sentiment scores, we fit separate models for each year, and plot estimates over time to understand how stylistic evolved. We fitted robust linear regressions with Huber loss to account for the skewed distribution of these features, with a categorical predictor coding for which agency has produced a given

tweet as predictor and the value of the target feature as the outcome variable. As for sentiment models, the sign of significant regression coefficient provides information on whether a given agency tends to score higher or lower than the EC on the target metric, and their values provide information on the magnitude of the effect. A positive coefficient for the ECB on sentence length metrics, for example, would suggest that the ECB produces, on average, longer sentences than the EC. Outcome variables were standardized before model fitting, to guarantee comparability across features.

The results are displayed in Fig. 8. Interestingly, overall reading complexity has decreased for the EC than for all agencies (except one). While EC tweets have evolved to be on average longer, they have also become less complex at the sentence level compared to those from all other agencies. As a result, the EC now produces
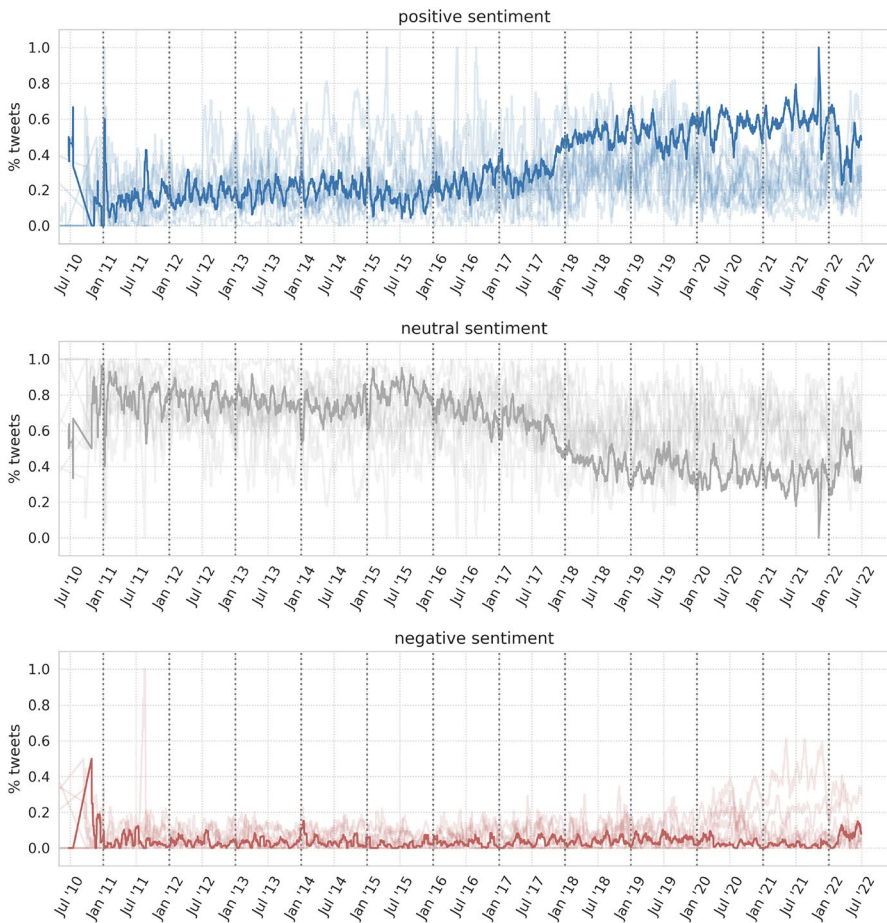

Fig. 5 Proportion of tweets with positive, neutral and negative sentiment over time. The dark line represents data for the EC. Lighter lines represent data for all other reference agencies. All time series are smoothed using rolling averages over 21 days
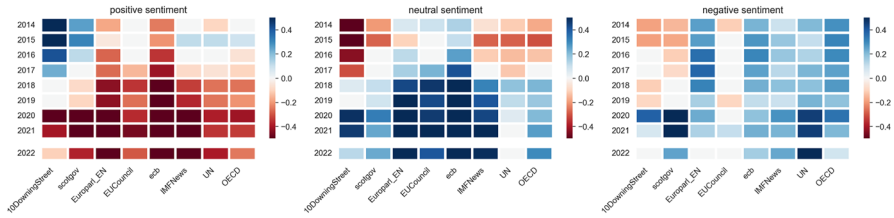
Fig. 6 Regression coefficients for pairwise comparisons between EC tweets and other agencies over time. Negative values (red) indicate cases where the EC tends to score higher on a given feature than the agency to which it is compared (x-axis). Positive values (blue) indicate cases where the EC tends to score lower on a given feature than the agency to which it is compared. Values are set to zero if the coefficient is not significant ($p > 0.05$)
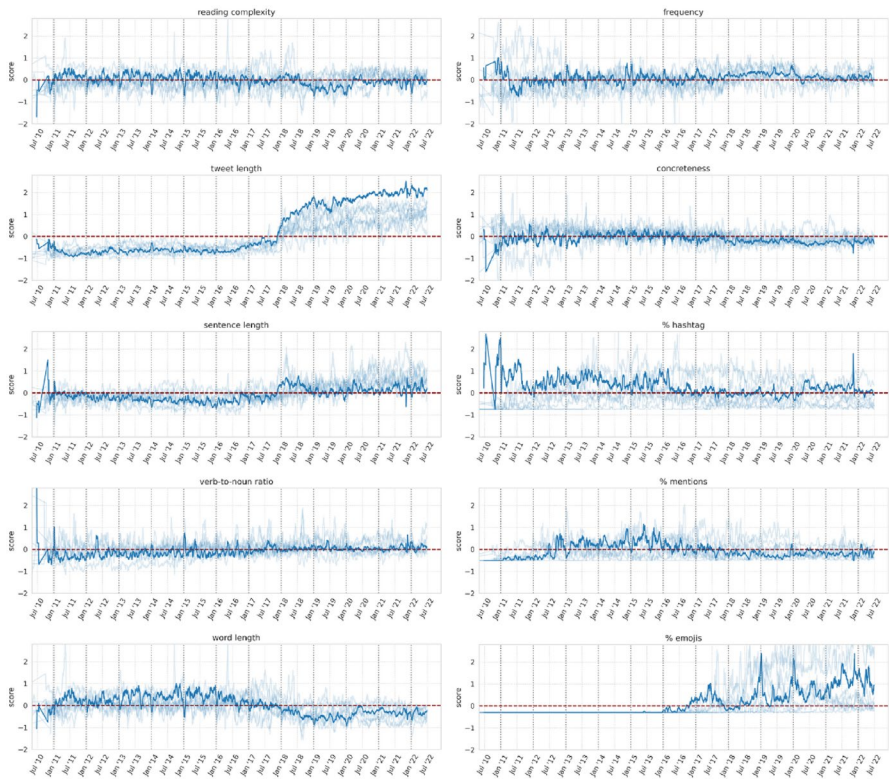
Fig. 7 Style features over time. Dark lines represent data for the EC. Lighter lines represent data for all other reference agencies. All time series are smoothed (rolling average over 21 days). Scores are normalized to zero mean and unit variance, scaling across all reference institutions for comparability. Frequency and concreteness indicate average word-level frequency, extracted from SubtlexUS [24], and average word-level concreteness for each tweet, extracted from [32]

Fig. 8 Robust regression coefficients showing group differences between each reference agency and the European Commission for a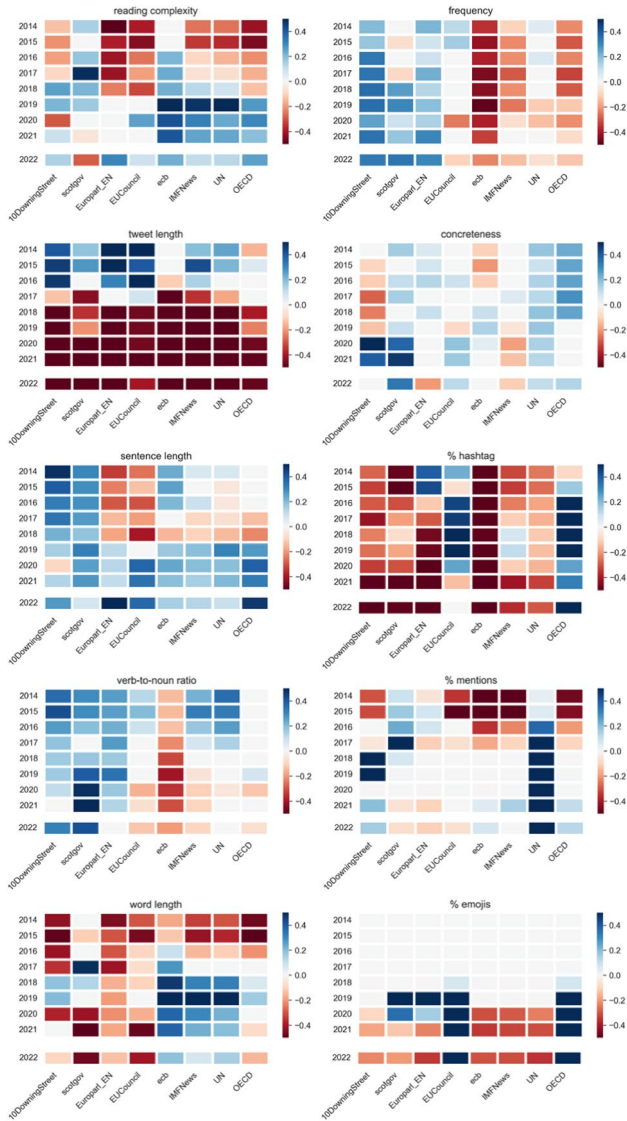ll platform-specific communication features. The analysis focuses on data from 2014 to 2022, as for some agencies, tweets are only available from 2014 on. Positive coefficients indicate that the target agency scores on average higher than the EC on that descriptive. Parameters are set to zero if the p-value is above 0.05

tweets with average reading complexity lower than or comparable to other agencies. The picture is more heterogeneous for verb-to-noun ratio and word-level complexity. Verb-to-noun ratio is higher than technocratic institutions like the ECB and IMF, but lower than national governments and the European Parliament. Action-orientedness

has evolved to become higher than technocratic agencies and other international agencies, but lower than national governments. On the other hand, EC communication is less lexically complex than technocratic agencies, but more lexically complex and less concrete than most other institutions, especially national governments. Finally, estimates for Twitter-specific features shows that the EC uses significantly more emojis than almost all other agencies, and more hashtags.

Overall, these patterns show a marked improvement in accessibility of EC communication over time. EC tweets started off being complex in comparison to all other agencies, and evolved towards a style which is simpler than all other agencies on many metrics. Yet, for lexical accessibility metrics and concreteness, EC communication is still consistently more demanding than that of, for example, national governments, a pattern observed in previous work focusing on other media [12], and replicating the finding that EU supranational actors in general have a less accessible style than national governments [5].

Predictors of engagement across content and style

In the previous analyses, we showed that the style of EC communication has evolved to become more aligned with topics that engage the general public, and stylistically more accessible, engaging, and more multimodal, potentially a strategy to increase self-legitimation and support. However, we did not yet provide direct evidence of how each of the variables analyzed directly relate to engagement. In the following, we address our fourth and final research question and test whether there is evidence in our data that topics on which EC communication has focused increasingly, and features of styles (traditionally related to accessibility and engagement) analyzed in the previous section do, indeed, elicit more engagements (retweets, likes, quote-tweets, and responses). The purpose of this analysis is two-fold: first, we aim to corroborate the hypothesis that observed changes in topics and style may be related to engagement; second, we provide additional, data-driven insights on drivers of engagement specific to the EC's Twitter public.

To do so, we trained an XGBoost model [38] predicting engagement values based on topic descriptors and style descriptors extracted for previous analyses. Engagement for any given tweet is computed as the sum of all engagement counts (likes, retweets, quote-tweets, replies) normalized (i.e., divided) by the number of users following the EC's Twitter account on the day where the tweet is posted. We normalized by the number of followers, as normalizing by the number of impressions was not possible, since Twitter's Academic API did not provide access to this information. Follower counts for each day were retrieved using open source functions from the following repository: https://github.com/ChRauh/PastTwitter, which crawl Wayback Machine (https://archive.org/web/) to retrieve follower counts at different time points. The earliest available data were from February 16th, 2013, and the latest available data within our time frame of interest were from June 15th, 2022. We used linear interpolation to infer follower counts for all days between the earliest and the
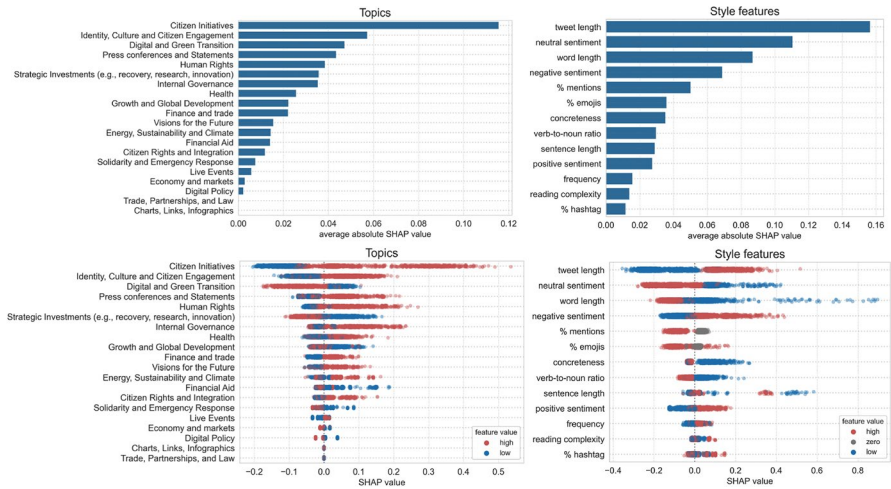
Fig. 9 SHAP values for all predictors included in the XGBoost model predicting engagement data. Plots in the top row display mean absolute SHAP values for each predictor, a proxy for feature importance. Plots in the bottom row display raw SHAP values for all examples in the test set. For plots in the bottom row, high feature values corresponding to positive values on the x-axis indicate that higher feature values are associated with higher values of the outcome variable (the two variables are positively associated). Viceversa, low feature values in the negative half of the x-axis indicate that lower feature values are associated with lower values of the outcome variable (the two variables are negatively associated). For some variables, relationships are more idiosyncratic

latest available date for which follower counts could not be retrieved. Tweets outside this temporal range were not used for the predictive modeling analysis.

While associations between features and engagement could be tested using statistical significance testing, we chose a predictive setup using tree-based methods for a number of reasons. First, predictive out-of-sample testing yields higher generalizability [39]; secondly, tree-based methods are robust to highly heterogeneous predictors; thirdly, these methods are agnostic to whether relations between features and outcomes are linear, and to the distribution of input features. We performed random grid search to tune hyperparameters, and converged on a model that yields $R^2 = 0.09$, and rank correlation between true and predicted engagement values of 0.39. Figure 9 displays feature importance values and SHAP values for the resulting model [40]. SHAP values are coefficients that make it possible to understand how a given predictor influences model outcomes (positively or negatively).

We observe that content related to identity and citizen engagement, citizen initiatives, human rights, and health is positively associated with engagement. These are all topics on which EC Twitter communication has placed increasingly more focus. Topics having a positive impact on engagement also include, perhaps surprisingly, internal governance and finance and trade, topics on which the EC has focused increasingly less over time. On the other hand, economy-related tweets (a

closely related topic from which the EC has progressively drifted away) seem to be negatively associated with engagement. Interestingly, topics related to digital and environmental policy are not positively related to engagement in our model.

Concerning style, we observe that tweets with lower lexical complexity, and tweets with lower overall reading complexity, as well as non-neutral tweets, elicit higher engagement, suggesting that changes in EC style may lead to more engagement. Interestingly, higher verb-to-noun ratio and higher concreteness yield less engagement, while longer tweets elicit more engagement—contrary to what we expected from previous literature on other media.

These changes in topics and style of EC communication over time could be due to strategic efforts to increase public engagement, but future research should test this link explicitly. Furthermore, it is important to highlight that increasing alignment with drivers of engagement might reflect efforts to produce content that aligns with preferences of the EC's *own* audience, while not necessarily aiming or succeeding at increasing engagement with the general public.

Discussion

We leveraged advanced natural language process methods and publicly available Twitter data to describe how the content and style of the EC's Twitter communication has evolved between 2010 and 2022. As well as providing a quantitative overview of how the focus of EC communication on different areas of policy has changed and how its style has evolved, we analyzed whether the EC has aligned with engagement patterns, devoting more attention to topics that generated higher engagement in the past and enhancing aspects of its messaging style that tend to yield higher engagement.

In analyzing dynamic changes in the content of EC Twitter communication, we observed a progressive detachment from topics that overlap with the focus of "technocratic" institutions included in our sample (e.g., economy, finance, and internal governance) and increased attention to topics related to social policy, environmental and digital policy, as well as identity and citizen participation. As a result of these changes, the EC has increasingly represented itself in ways that make it highly distinct from technocratic institutions, and generally more unique among comparable agencies. Furthermore, several of the topics on which the EC has focused more in recent times are positively associated with audience engagement. Additionally, the EC's communication has increasingly aligned with the preferences of its online audience, with alignment between relative volume of each topic and relative number of engagements per topic at previous time points increasing steadily over time and at a faster rate than other institutions.

The style of EC communication has also evolved significantly over the years. We observed a notable shift in the sentiment of EC tweets. While tweets with neutral sentiment were prevalent in earlier years, the proportion of tweets with neutral sentiment has notably decreased over time, and the proportion of tweets with positive sentiment has increased steeply. Compared to other agencies, the EC currently produces significantly fewer tweets with either neutral or negative sentiment, and significantly more tweets with positive sentiment. Non-neutral sentiment has been previously identified as an important driver of online engagement and a factor influencing the potential of online content to

spread [21–23]. These findings are in line with the results of our predictive modeling analysis, where we observe a negative association between neutral sentiment and overall number of engagements. Based on this evidence, this radical shift in the sentiment distribution of EC tweets—which may reflect an intentional change in communication strategy—can be considered an important step towards developing a messaging style that favors engagement and increases public support. Note, however, that our engagement metrics capture engagement within the EC's own public, which might not be representative of the general public. Efforts to increase engagement by modulating content and style of tweets may thus not necessarily be targeted to and/or successful at increasing general engagement and support, but rather limited to increasing engagement within this self-selected readership or followers with similar characteristics.

The EC's messaging style has also evolved in terms of overall readability and complexity. We observed that, while the overall length of tweets has increased—with the EC currently producing significantly longer tweets compared to all other institutions in our sample—EC messaging has become lexically simpler, more action-oriented, more readable, and more multimodal over time. These stylistic features are commonly associated with more accessible messaging, a prerequisite for successful communication [5, 11, 12]. As a result, the EC currently scores better than all other reference institutions on compound readability indices, and better than technocratic institutions on several fine-grained indicators of stylistic complexity and accessibility. However, we observe that the EC's messaging style is still more lexically complex, less action-oriented and less concrete than other institutions in our sample, especially national governments. These findings are partly in line with recent research observing that the EC's messaging style on traditional media seems to be less accessible than that of a range of comparable institutions [12].

In line with hypotheses from previous studies, our predictive modeling analyses provide evidence of negative associations between lexical complexity and engagement, and between overall reading complexity and engagement. Interestingly, however, we also observe that less action-oriented, less concrete, and longer tweets tend to yield higher engagement. These findings are in contrast with the assumptions made in previous studies (e.g., [12]). Preferences for longer, less concrete and less action-oriented text might be specific to the EC's Twitter audience, or specific to Twitter communication. While our results do not support qualified inference on the drivers of these audience-specific effects, they raise novel questions on the relations between text descriptors traditionally associated with accessibility and empirical engagement patterns across multiple public communication channels and sociodemographic subgroups to be addressed in future research.

Finally, we observed that 2017 seems to mark a key moment of transition for EC communication both in relation to the evolution of its topics and in relation to the evolution of its messaging style. We speculate that this may be a reflection of the politicization shock represented by the Brexit referendum, a hypothesis that warrants further investigation in future research.

Future directions

Our study provides a comprehensive overview of how the topics and style of the EC's communication on Twitter have evolved between 2010 (the year where the EC's official Twitter account was created) and present days. Previous literature has underlined how social media can be an important strategic tool to increase perceived legitimacy and citizen engagement in contexts of high politicization [1–3], especially for supranational bodies with indirect electoral accountability like the European Commission [5]. Overall, our analyses show that the EC has evolved a communication strategy that better aligns with its audience. This suggests that the EC may have become increasingly more successful in exploiting the potential of social media to promote legitimacy and public support. However, empirical validation of the impact of EC social media communication on perceived legitimacy and public support is needed to corroborate this hypothesis.

To this end, future directions include more hypothesis-driven and experimental studies combining Twitter data with public opinion data sources, to directly investigate relations between changes in online communication and empirical metrics of perceived legitimacy and support. In addition, cross-referencing Twitter data with external policy documents and data from other communication channels (e.g., press releases) may help elucidate whether changes observed in the EC's online identity are merely reflective of evolving policy priorities, or they result from effort to evolve more efficient communication and digital diplomacy strategies.

Finally, there are caveats concerning our findings on predictive relations between linguistic features and engagement data. More adequate indicators of tweets' engagingness could be constructed by normalizing the raw number of engagements (e.g., likes or retweets) by the number of impressions, that is, the number of times the tweet appears on users' timelines. Unfortunately, impression counts are not available through the Twitter API, and engagement data were normalized by follower counts, which is not the optimal proxy for engagement rates.

Conclusion

Even on the relatively short time scale offered by Twitter data, we were able to describe important dynamic trends that point to the evolving nature of EC online communication and its emerging digital identity. Our work represents a foundational step towards future hypothesis-driven research, and a demonstration of how state-of-the-art NLP tools can be used to analyze the complex dynamics characterizing the nature of supranational institutions' digital communication.

Further methodological details

Text preprocessing

The preprocessing pipeline for text involved the following steps: (1) removing retweets and mentions, respectively identified as tweets beginning with the string "RT" (as per Twitter API convention), and tweets beginning with the symbol "@". Links (identified by the presence of the substring "http") were stripped from the tweet's text, but tagged tweets by whether they originally contained a link as one of the features of interest for the stylistic analysis. We further normalized transcription for non-ASCII encodings (e.g., &amp -¿ &). We removed all tweets that were not tagged as English by the Twitter API (as indicated by the "lang" field in the API response), and all tweets that after pre-processing included no residual characters (e.g., link-only tweets). As some of our pipelines involve transformer models—which model both lexical and non-lexical components of linguistic input, have a rich token vocabulary, and are robust to misspelling and spelling variability—tweets did not undergo any further preprocessing. Some of the accounts included in the analysis occasionally tweet in languages other than English. This is also the case of the European Commission's Twitter account—where a small percentage of tweets are in other languages, but for all non-English tweets, a corresponding tweet in English is available. As mentioned, the present study only focuses on English tweets.

Engagement metrics

The following fields were natively available from the Twitter API: number of retweets ("retweet_count"), likes ("like_count"), quotes ("quote_count"), and replies ("reply_count"). To extract a simpler summary measure of engagement, we compute the sum of all the four engagement metrics (henceforth: "engagement"). Note that the Twitter API does not provide access to the number of times a given tweet was visualized, a metric which is required to compute the rate of engagement, a more stable measure of engagement. Since visualization counts cannot be accessed, we normalize engagement counts by the number of followers of the EC account on the day where the tweet was produced.

Data splits

70% of the tweets from the European Commission's official account were assigned to the training set for topic models, while the remaining tweets were evenly divided into a 15% validation and a 15% test set, used for early stopping and model selection. Compared to traditional approaches to topic modeling, where models are trained and evaluated on the same data, this procedure ensures that estimates of model performance are more robust and generalizable,

and provide a better prior on the model's potential to accurately describe new tweets, not included in the current sample. The same splits were used to estimate XGBoost models.

Contextualized topic models

Contextualized topic models have been shown to improve intra-topic coherence and inter-topic distinctness compared to both bag-of-words and more traditional neural topic modeling methods, by supplementing bag-of-words representations with representations from transformer models [20]. To estimate the best parameters for our topic models, we fitted a family of models varying in parameter configurations. We parametrically varied the number of target topics (varying between 10 and 100), learning rate (2e-2 to 2e-5), batch size (16, 64 tweets), and vocabulary size for the bag-of-words model (250, 500, 1000 words). We also varied the pretrained transformer model used to extract contextualized embeddings, experimenting with both a standard pretrained Distil-BERT model [9], a version of DistilBERT fine-tuned on masked language modeling using our own tweet corpus, a range of sentence transformers [41], and a model fine-tuned on generic topic extraction from Twitter content available on the HuggingFace Model Hub (https://huggingface.co/cardiffnlp/tweet-topic-21-multi). To select models that not only describe the dataset used for the present study, but also have the highest chance of generalizing to future tweets, we extracted a training and a validation set from the EC's tweets (covering 70%, 15% of the tweets respectively), trained topic models on the training set, and used the validation set as an evaluation set for early stopping. The best model was selected on the basis of topic coherence on a test set that included both EC test set tweets, and test sets from all other agencies. Coherence was operationalized as normalized pointwise mutual information (NPMI) between topic-defining words [42], a commonly used metric in the field. By estimating the model on European Commission's tweets and evaluating on test sets from all agencies, we make sure that the selected model: (a) is specifically tailored to identifying semantic attractors in the EC's Twitter communication; (b) also identifies a set of topics that are also represented in tweets from other agencies, allowing for comparisons. As neural topic models can display unstable estimates [43], we fit each model for 5 distinct runs. The best performing model was a pre-trained DistilBERT model with 20 target topics and a 500-word vocabulary (10-word NPMI = 0.064 on the unseen test set). The resulting model was used to extract per-tweet scores for each topic. These are used for analyses and visualizations focusing on the temporal evolution of topic volumes over time, as well as for predictive modeling analyses investigating the relation between linguistic features and engagement.

Aggregating style features using principal component analysis

We used Python packages TextDescriptives [33] and pliers [34] to extract all style metrics analysed in the paper. For features (e.g., readability, word length) for which multiple indicators are available on TextDescriptives, we performed principal component analysis to reduce overlapping feature sets to a single feature.

An aggregate readability index was extracted by performing a principal component analysis on the 7 readability indices available through the TextDescriptives package used for feature extraction, which included: the Flesch Reading Ease index, the Flesch-Kincaid Grade, the Gunning Fog index, the Automated Readability index, the Coleman-Liau index, LIX, and RIX. An in-depth description of each metric is provided in the package documentation, available at: https://hlasse.github.io/TextDescriptives/readability.html. PCA yielded a first component which explained over 80% of the variance, which was used as aggregate readability index for all analyses. Loadings were negative for all readability indices except for the Flesch reading ease, reflecting the fact that this index quantifies ease of reading, while all other indices quantify reading complexity. The resulting index is therefore to be interpreted as an indicator of reading complexity.

The sentence length variable used in the style analyses also results from a PCA conducted on the two variables available through TextDescriptives which quantify sentence length in terms of number of words and number of characters. For simplicity, we reduced them to the first principal component of a PCA estimated on both variables, which captured around 97% of the variance. Loadings for both variables were positive, meaning that the final aggregate variable displays higher values for longer sentences.

The same procedure was used to aggregate the two measures of word length (in characters and in syllables) available through the TextDescriptives package (with the first principal component capturing 92% of the variance and both loadings being positive) and to aggregate the three measures of overall text length (number of words, number of characters, and number of sentences), where the first principal component captured around 78% of the variance and all loadings were positive.

XGBoost model

To analyze the relation between topic and stylistic features and engagement, we fitted an XGBoost model, where we used the sum of all engagement types (likes, retweets, quotes, and responses) as the outcome variable, and all topic and style predictors mentioned in the respective sections as input features. We split EC tweets into 70/15/15 train/validation/test splits (see "Data splits" subsection), and used the validation set to perform grid-based hyperparameter

optimization. We used Tweedie loss [44] to best fit the highly skewed distribution followed by engagement data. The best model (used for the analyses presented in the manuscript) had a rank correlation of 0.39 with engagement data on the test set, and $R^2 = 0.09$ on the test set. We performed hyperparameter search over the following parameter grid, iterating over 2000 combinations:

- learning_rate: [2e-5, 2e-3, 2e-2, 2e-1]
- min_child_weight: [1, 5, 10, 50]
- gamma: [0,.5, 1., 2.]
- subsample: [.6,.8, 1]
- colsample_bytree: [.3,.5,.7, 1]
- max_depth: [2, 3, 5, 10, 20]
- reg_alpha:[0,.1, 5.]
- reg_lambda: [0,.1, 1.]
- n_estimators: [1, 5, 10, 30, 50]
- tweedie_variance_power: [1.01, 1.3, 1.6, 1.8, 1.99]

The best model (selected on the basis of its root mean squared error on the validation set) had the following parameters: learning_rate = 0.2; min_child_weight = 10; gamma = 1.0; subsample = 0.8; colsample_bytree = 1; max_depth = 3; reg_alpha = 0.1; reg_lambda = 0.1; n_estimators = 50; tweedie_variance_power = 1.3.

Appendix

See Figs. 10, 11, 12, 13 and Tables 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17.
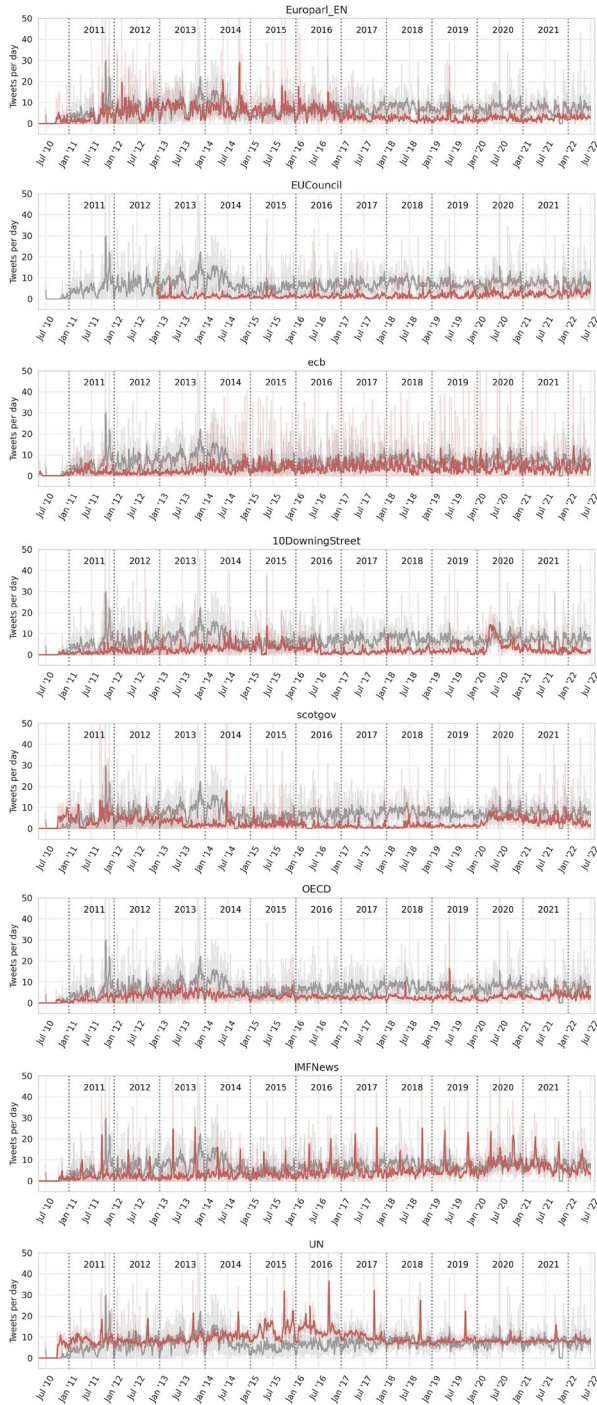
Fig. 10 Tweet volume over time for each agency in the database (red lines) vs EC (gray lines)
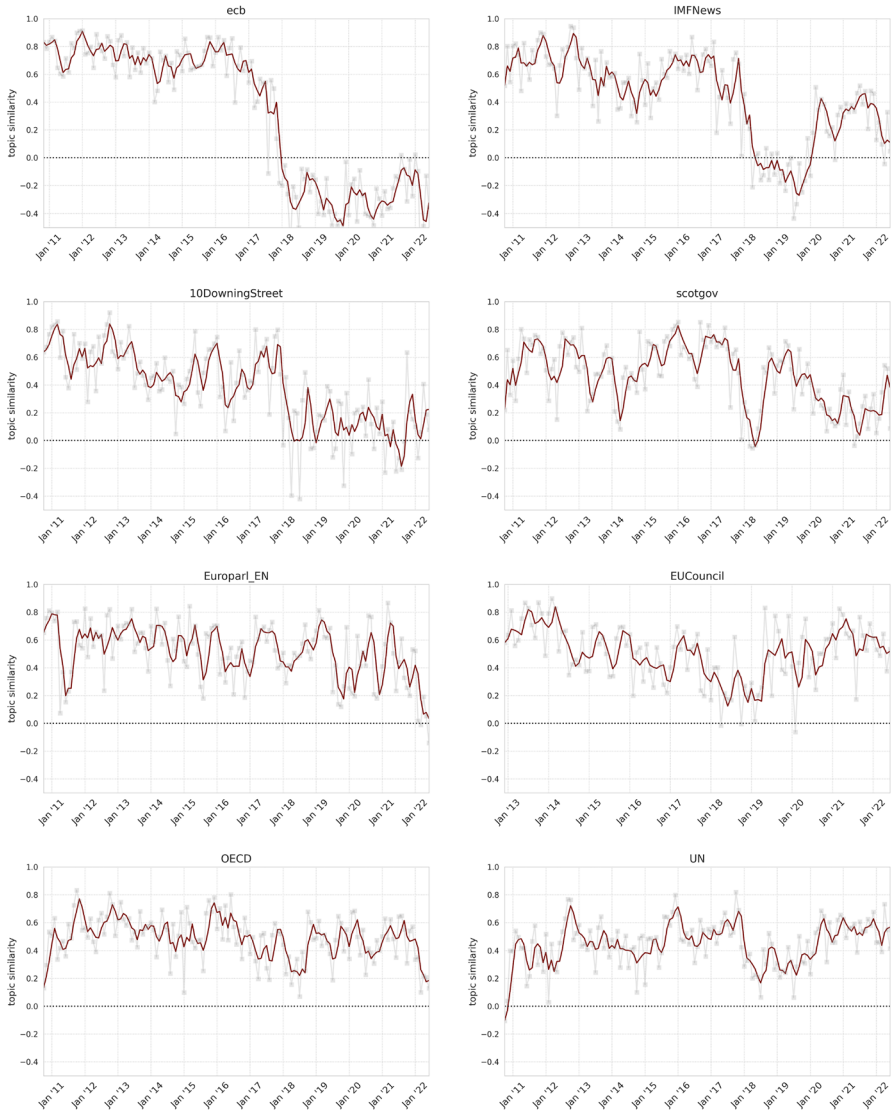
Fig. 11 Topic-based similarity (rank correlation between topic volumes for each of the 20 topics) per month between the EC's Twitter account and other official accounts
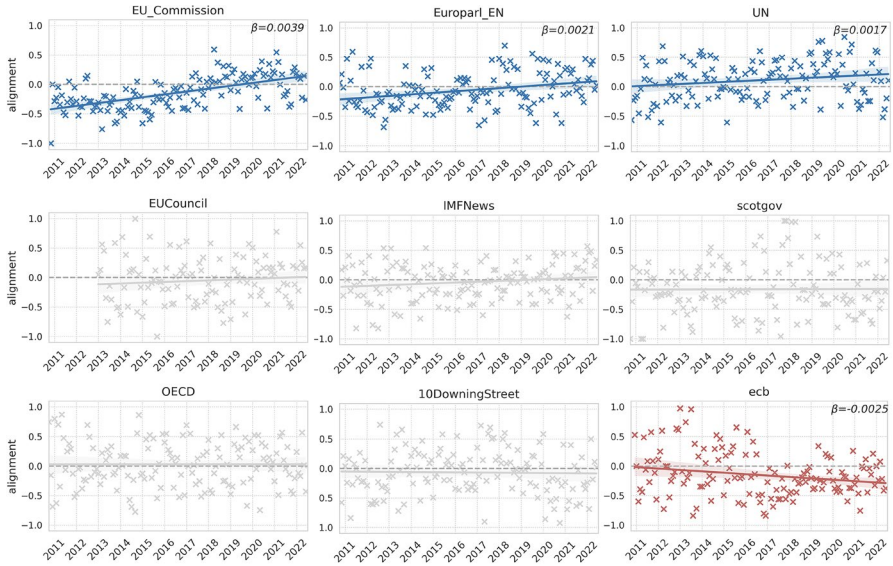
Fig. 12 Alignment between topic volume and engagement over time. Values on the y-axis are alignment indices, operationalized as correlations between the rank of a given topic in the tweet volume distribution at month m, and its rank in the engagement distribution at month m − 1. Positive regression coefficients indicate increasing volume for topics that previously generated high engagement. No regression coefficient (grey) indicates that the coefficient is not significant (at $p < 0.05$)

Fig. 13 PCA loadings for principal component analyses on style metrics

Table 3 Parameter estimates for models quantifying difference in positive sentiment over time between EC and other institutions

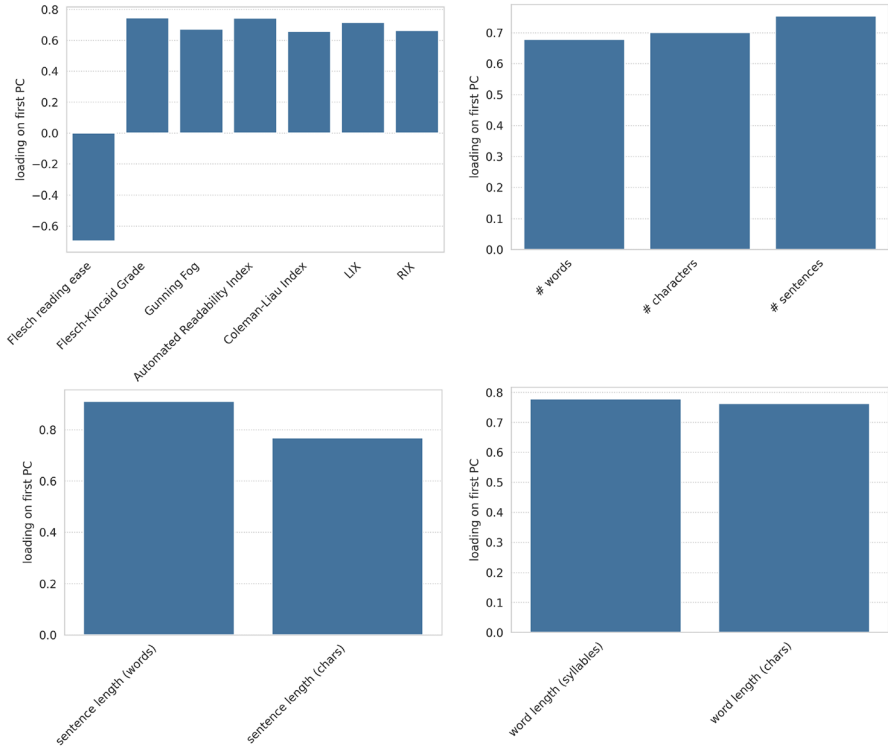| Predictor year	10Downing-Street	EUCouncil	Euro-parl_EN	IMFNews	OECD	UN	ecb	scotgov
2014	0.742302	−0.000000	−0.177910	−0.000000	−0.000000	−0.094463	−0.242501	0.256400
2015	0.774890	0.000000	−0.044413	0.116275	0.087254	0.115703	−0.208676	0.306299
2016	0.417574	0.000000	−0.276311	−0.000000	−0.058634	−0.000000	−0.348909	0.117281
2017	0.228336	−0.150019	−0.268287	−0.082512	−0.097864	−0.062412	−0.416044	0.000000
2018	−0.000000	−0.347777	−0.435507	−0.334934	−0.255103	−0.255453	−0.551575	−0.083649
2019	0.000000	−0.258485	−0.430100	−0.372094	−0.209596	−0.251723	−0.517380	−0.096206
2020	−0.579448	−0.363596	−0.566185	−0.558066	−0.413358	−0.386850	−0.669763	−0.674652
2021	−0.404693	−0.461194	−0.501700	−0.524626	−0.337621	−0.350975	−0.615560	−0.624850
2022	−0.105625	−0.295278	−0.527459	−0.507923	−0.249236	−0.395202	−0.531762	−0.382085

Table 4 Parameter estimates for models quantifying difference in neutral sentiment over time between EC and other institutions

Predictor year	10Down-ingStreet	EUCouncil	Euro-parl_EN	IMFNews	OECD	UN	ecb	scotgov
2014	−0.596291	0.095279	0.066341	−0.062898	−0.174737	−0.116423	0.142110	−0.199665
2015	−0.693793	−0.000000	−0.105048	−0.278291	−0.306617	−0.284579	0.107391	−0.279781
2016	−0.441126	−0.000000	0.118521	−0.113553	−0.129560	−0.143003	0.256853	−0.000000
2017	−0.309506	0.215185	0.152311	−0.000000	−0.000000	−0.123578	0.416753	0.000000
2018	0.054780	0.470562	0.437162	0.310288	0.212649	0.184647	0.676048	0.107427
2019	−0.000000	0.433549	0.513603	0.409249	0.168003	0.115709	0.634890	0.082113
2020	0.483100	0.463786	0.553839	0.439077	0.206584	0.051987	0.713406	0.322124
2021	0.460976	0.537431	0.549058	0.507075	0.259759	0.000000	0.643612	0.240807
2022	0.124589	0.409733	0.710619	0.491066	0.293053	0.000000	0.605743	0.234460

Table 5 Parameter estimates for models quantifying difference in negative sentiment over time between EC and other institutions

Predictor year	10Downing-Street	EUCouncil	Euro-parl_EN	IMFNews	OECD	UN	ecb	scotgov
2014	−0.184862	−0.111129	0.257490	0.094564	0.284109	0.203319	0.216818	−0.150106
2015	−0.189569	−0.000000	0.259008	0.172305	0.305176	0.105891	0.274220	−0.178914
2016	−0.000000	−0.000000	0.354239	0.173523	0.296526	0.138297	0.301532	−0.112655
2017	−0.000000	−0.000000	0.373328	0.124808	0.252451	0.211114	0.253163	−0.082884
2018	−0.111897	0.000000	0.310792	0.188778	0.266152	0.188886	0.223884	−0.000000
2019	−0.092968	−0.095219	0.148888	0.100267	0.184695	0.217349	0.112862	0.000000
2020	0.384479	−0.000000	0.232494	0.280955	0.347332	0.453781	0.187325	0.643676
2021	0.072687	0.109127	0.141833	0.217776	0.207490	0.454262	0.223868	0.705140
2022	−0.000000	0.000000	0.000000	0.240204	0.082616	0.510912	0.154119	0.248534

Table 6 Parameter estimates for models quantifying difference in reading complexity over time between EC and other institutions

Predictor year	10Downing-Street	EUCouncil	Europarl_EN	IMFNews	OECD	UN	ecb	scotgov
2014	−0.133679	−0.391216	−0.577179	−0.251219	−0.421695	−0.251225	−0.000000	0.145389
2015	−0.212504	−0.442216	−0.386865	−0.340072	−0.455142	−0.365451	0.000000	0.000000
2016	−0.188994	−0.253268	−0.413440	−0.092881	−0.215237	−0.154318	0.214028	0.175939
2017	−0.090514	−0.189357	−0.414758	−0.059819	−0.161737	−0.070942	0.231044	0.489936
2018	0.252491	−0.318248	−0.238478	0.124987	−0.133884	0.088272	0.225898	0.211533
2019	0.213016	0.000000	−0.000000	0.479496	0.265976	0.510566	0.508090	0.165065
2020	−0.295032	0.252779	−0.000000	0.314681	0.297042	0.237197	0.414671	−0.000000
2021	0.098925	0.000000	−0.000000	0.243953	0.194262	0.238481	0.409717	−0.052007
2022	0.143841	0.090746	0.301706	0.112769	0.252401	0.124366	0.226267	−0.286766

Table 7 Parameter estimates for models quantifying difference in frequency over time between EC and other institutions

Predictor year	10Downing-Street	EUCouncil	Europarl_EN	IMFNews	OECD	UN	ecb	scotgov
2014	0.201682	0.117604	0.212846	−0.104445	−0.220847	0.048617	−0.354060	0.000000
2015	0.173715	0.155586	0.084789	−0.220464	−0.296125	−0.000000	−0.425687	−0.071573
2016	0.333805	0.000000	0.092614	−0.218406	−0.226828	0.059944	−0.385129	−0.000000
2017	0.341086	−0.000000	0.222964	−0.291458	−0.327856	0.000000	−0.455066	−0.086438
2018	0.356231	0.000000	0.132270	−0.215091	−0.291642	0.000000	−0.440450	0.260301
2019	0.352216	0.000000	0.193489	−0.248411	−0.117591	−0.114061	−0.490835	0.297684
2020	0.249338	−0.245007	0.144436	−0.072974	−0.217829	−0.138231	−0.384137	0.092943
2021	0.345791	0.000000	0.303258	0.000000	−0.081011	0.000000	−0.328633	0.172691
2022	0.328828	−0.101685	0.298334	−0.137847	−0.128561	−0.067011	−0.227879	0.336343

Table 8 Parameter estimates for models quantifying difference in tweet length over time between EC and other institutions

Predictor year	10Downing-Street	EUCouncil	Europarl_EN	IMFNews	OECD	UN	ecb	scotgov
2014	0.403937	0.783885	0.605402	0.198176	−0.156036	0.244780	−0.000000	0.180070
2015	0.457004	0.400880	0.516862	0.423239	0.059645	0.212839	−0.000000	0.272917
2016	0.465165	0.607554	0.311074	0.144022	0.000000	0.000000	−0.115665	−0.000000
2017	−0.132375	0.085046	−0.000000	−0.354262	−0.000000	−0.179535	−0.707712	−0.439473
2018	−0.733482	−0.470779	−0.648209	−0.734675	−0.394872	−0.711822	−1.060871	−0.351804
2019	−0.609377	−0.471585	−0.766768	−0.770830	−0.237211	−0.759043	−1.078206	−0.213303
2020	−1.131101	−0.466002	−0.934113	−0.902522	−0.468914	−0.616573	−1.018840	−0.519311
2021	−1.261785	−0.540575	−1.163204	−0.938495	−0.488566	−0.654309	−0.911653	−0.739532
2022	−0.827648	−0.405000	−1.345319	−0.945173	−0.484087	−0.674337	−0.741815	−0.959192

Table 9 Parameter estimates for models quantifying difference in concreteness over time between EC and other institutions

Predictor year	10Down-ingStreet	EUCouncil	Euro-parl_EN	IMFNews	OECD	UN	ecb	scotgov
2014	0.000000	0.000000	0.061661	−0.000000	0.231048	0.178079	−0.107268	0.153952
2015	−0.100555	−0.000000	0.080639	−0.000000	0.244873	0.097422	−0.206876	0.000000
2016	−0.117020	0.094268	0.131298	0.000000	0.247177	0.136793	−0.110579	0.107088
2017	−0.278013	−0.000000	0.000000	0.058561	0.283511	0.122700	−0.000000	0.148600
2018	−0.239547	0.000000	0.105073	−0.000000	0.243698	0.180802	0.103784	0.000000
2019	−0.132416	−0.083115	0.000000	−0.053776	−0.000000	0.173311	0.138115	0.101880
2020	0.759610	0.182725	−0.000000	−0.188745	−0.000000	0.141037	0.046844	0.372228
2021	0.391572	0.166353	−0.000000	−0.136392	0.000000	0.126388	0.000000	0.464881
2022	−0.000000	0.118109	−0.200522	−0.110572	0.134855	0.109955	0.000000	0.335765

Table 10 Parameter estimates for models quantifying difference in sentence length over time between EC and other institutions

Predictor year	10Down-ingStreet	EUCouncil	Euro-parl_EN	IMFNews	OECD	UN	ecb	scotgov
2014	0.482452	−0.255335	−0.337233	0.067335	−0.000000	0.067296	0.231045	0.301760
2015	0.397463	−0.135864	−0.234528	0.000000	0.000000	−0.047698	0.250377	0.287875
2016	0.321010	−0.304434	−0.293636	0.100457	0.000000	−0.046239	0.210518	0.281474
2017	0.341903	−0.192170	−0.147460	−0.091304	−0.136925	−0.069567	−0.000000	0.264458
2018	0.218895	−0.401491	−0.181733	−0.103041	−0.224188	−0.141175	−0.150482	0.180927
2019	0.151703	−0.000000	0.119937	0.194343	0.268012	0.309524	0.171490	0.318923
2020	−0.090568	0.371360	0.128386	0.197346	0.384662	0.194424	0.240694	0.255700
2021	0.134361	0.345176	0.109679	0.148722	0.327841	0.205114	0.194757	0.272582
2022	0.268408	0.368108	0.508351	0.132562	0.468113	0.102100	0.150767	0.079369

Table 11 Parameter estimates for models quantifying difference in % hashtags over time between EC and other institutions

Pre-dictor year	10Downing-Street	EUCouncil	Euro-parl_EN	IMFNews	OECD	UN	ecb	scotgov
2014	−0.275508	0.227913	0.393048	−0.316981	−0.086613	−0.249969	−0.700563	−0.481337
2015	−0.331498	−0.069958	0.426579	−0.337663	0.165023	−0.265125	−0.742017	−0.502087
2016	−0.292308	0.442331	−0.117337	−0.238991	0.625784	−0.181755	−0.677340	−0.308337
2017	−0.420620	0.456352	−0.295955	−0.112608	0.932362	−0.159402	−0.553979	−0.190308
2018	−0.238078	0.688381	−0.465574	0.073653	0.853903	−0.120983	−0.489913	−0.076771
2019	−0.258136	0.848200	−0.476246	0.110737	0.681236	−0.137506	−0.492262	−0.178462
2020	−0.335083	0.242597	−0.603158	−0.110575	0.248089	−0.156128	−0.657335	−0.303981
2021	−0.639808	−0.138481	−0.803049	−0.412817	0.316496	−0.328695	−0.806162	−0.559509
2022	−0.597923	0.000000	−0.703166	−0.367146	0.729405	−0.294256	−0.630523	−0.491315

Table 12 Parameter estimates for models quantifying difference in verb-to-noun ratio over time between EC and other institutions

Predictor year	10Downing-Street	EUCouncil	Europarl_EN	IMFNews	OECD	UN	ecb	scotgov
2014	0.360395	0.125379	0.259643	0.221424	0.000000	0.355721	−0.143442	0.289988
2015	0.423307	0.201210	0.254097	0.321491	0.000000	0.324811	−0.145165	0.291245
2016	0.262601	0.097455	0.226108	0.175507	−0.000000	0.259667	−0.133774	0.197245
2017	0.179893	0.000000	0.263618	0.076239	−0.000000	0.143888	−0.222397	−0.000000
2018	0.160793	0.000000	0.165631	−0.000000	−0.000000	0.000000	−0.319551	0.173323
2019	0.160782	0.000000	0.345186	−0.057175	0.077819	−0.000000	−0.406764	0.397048
2020	0.112415	−0.156912	0.170598	−0.107126	−0.122856	−0.048788	−0.350197	0.480776
2021	−0.000000	−0.067154	0.134804	−0.100800	0.000000	−0.000000	−0.306959	0.587753
2022	0.313002	−0.121803	0.000000	−0.095009	−0.069023	−0.000000	−0.198561	0.415191

Table 13 Parameter estimates for models quantifying difference in % mentions over time between EC and other institutions

Predictor year	10DowningStreet	EUCouncil	Europarl_EN	IMFNews	OECD	UN	ecb	scotgov
2014	−3.031147e−01	−0.348412	−5.818915e−02	−0.516476	−4.716473e−01	0.036458	−6.550390e−01	0.097651
2015	−3.127137e−01	−0.498148	7.864559e−02	−0.576129	−4.458394e−01	0.051576	−6.714202e−01	0.166705
2016	−0.000000e+00	−0.000000	8.033036e−02	−0.226655	−2.019101e−01	0.369260	−3.552900e−01	0.234238
2017	−4.852614e−02	−0.068946	−1.128384e−01	−0.097993	−9.768958e−02	0.603762	−1.742412e−01	0.995383
2018	1.031250e+00	0.000000	−1.884717e−10	−0.000000	0.000000e+00	0.891892	−2.508851e−10	0.073939
2019	1.292411e+00	0.001685	−1.708631e−03	0.002572	4.284825e−03	0.916265	−1.631533e−03	0.005229
2020	−1.810379e−07	−0.000000	−4.912128e−07	−0.000000	2.648499e−07	1.000000	−5.373879e−07	−0.000000
2021	1.952203e−01	−0.000000	−1.016090e−01	0.146511	1.131916e−01	0.760601	4.833290e−02	−0.062371
2022	1.538974e−01	−0.069944	−1.076687e−01	0.000000	1.234301e−01	0.696032	7.182680e−02	−0.073375

Table 14 Parameter estimates for models quantifying difference in word length over time between EC and other institutions

Predictor year	10Downing-Street	EUCouncil	Europarl_EN	IMFNews	OECD	UN	ecb	scotgov
2014	−0.428299	−0.298314	−0.450741	−0.329486	−0.469654	−0.297429	−0.177320	−0.000000
2015	−0.491544	−0.458745	−0.298670	−0.420560	−0.517768	−0.372182	−0.123662	−0.111055
2016	−0.418668	−0.099066	−0.298219	−0.149695	−0.212448	−0.113088	0.102041	0.000000
2017	−0.355152	−0.076150	−0.412581	0.000000	−0.000000	0.000000	0.274788	0.484841
2018	0.183711	−0.105124	−0.160592	0.313331	0.063208	0.317437	0.468218	0.133424
2019	0.202232	−0.000000	−0.198585	0.529341	0.163284	0.530947	0.588583	0.000000
2020	−0.379753	−0.162890	−0.215299	0.297414	0.000000	0.222093	0.377696	−0.417050
2021	−0.000000	−0.478738	−0.176199	0.224669	−0.082009	0.198939	0.378673	−0.516854
2022	−0.094178	−0.408200	−0.133994	0.073826	−0.152184	0.150067	0.184612	−0.469898

Table 15 Parameter estimates for models quantifying difference in % emojis over time between EC and other institutions

Predictor year	10DowningStreet	EUCouncil	Europarl_EN	IMFNews	OECD	UN	ecb	scotgov
2014	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2015	−3.6e−140	−3.6e−140	−3.6e−140	−3.6e−140	−8.4e−141	−3.6e−140	−3.6e−140	−3.6e−140
2016	−1.10e−45	−1.10e−45	−1.10e−45	−1.10e−45	−1.10e−45	−1.10e−45	−1.10e−45	−1.10e−45
2017	−1.56e−20	−1.56e−20	−1.56e−20	−1.56e−20	−1.56e−20	−1.56e−20	−1.568776e−20	−1.568776e−20
2018	−4.25e−12	7.07e−02	−2.52e−12	−5.18e−12	7.69e−02	−7.380648e−13	−5.36e−12	2.80e−12
2019	−1.52e−14	2.04e+00	1.12e+00	−4.36e−14	2.14e+00	−2.59e−14	−4.55e−14	9.21e−01
2020	−9.21e−02	1.75e+00	1.53e−01	−2.91e−01	1.22e+00	−2.40e−01	−3.08e−01	3.52e−01
2021	−1.81e−01	1.72e+00	−2.31e−01	−3.21e−01	1.38e+00	−3.10e−01	−3.42e−01	−1.21e−01
2022	−2.24e−01	1.08e+00	−3.42e−01	−3.11e−01	1.49e+00	−3.49e−01	−3.02e−01	−2.026e−01

Table 16 Parameter estimates for regression models analyzing alignment between distributions of topic volumes and engagement distribution

	Entity	Beta	p-value
EU_Commission	0.0039	0.000	
Europarl_EN	0.0021	0.001	
UN	0.0017	0.009	
EUCouncil	0.0008	0.484	
IMFNews	0.0008	0.175	
scotgov	0.0006	0.533	
OECD	−0.0001	0.939	
10DowningStreet	−0.0004	0.583	
ecb	−0.0025	0.001	

Table 17 Top scoring tweets per topic

Economy and markets	#Antitrust: Commission opens proceedings against #Motorola
	#Inflation in the #euro area has picked up, but underlying price pressures remain subdued
	July 2012, the Business Climate Indicator (BCI) for the #euro area decreased by 0.32 points to 1.27
	Commission proposes to cut #sulphur #dioxide #emissions from #shipping by 90%
	Euro area retail trade −0.4% in July 14 over June 14, +0.8% over July 13
	Proposal to protect the #euro and other currencies against #counterfeiting
	Antitrust: Commission confirms it is investigating #Deustch-eBahn
	Commission calls for action on #energy and #taxevasion and #fraud
	Antitrust: Commission market tests Standard and Poor's commitments on international securities identification numbers
	#Antitrust: Commission market tests Standard and Poor's commitments on international securities #identification numbers
Finance and trade	State aid: @EU_Commission approves investment aid for Marseille Provence airport and airport charges
	.@EU_Eurostat: Euro area annual #inflation down to 0.4% in July 2014
	May 2012: Economic sentiment falls sharply in both the #EU and the euro area
	.@EU_Eurostat: Euro area international #trade in goods surplus 0.9 bn euro
	.@EU_Eurostat: Euro area annual inflation down to 0.7%; EU down to 0.9%
	.@EU_Eurostat: Euro area annual inflation confirmed at −0.1% in March 2015
	Industrial producer prices up by 0.4% in both euro area and EU27
	December 2012 #Euro area annual #inflation stable at 2.2% #EU down to 2.3% via @EU_Eurostat
	Economic Sentiment declines moderately in both the EU and the #euro area
	Euro area annual inflation up to 0.0% in April 2015—flash estimate from @EU_Eurostat
Growth and Global Development	LNG to ensure diversity, reliability and competitiveness
	Making the most of the Data-Driven Economy
	Study calls for stronger focus on #IT and entrepreneurial skills in schools
	Internal #security: What progress has been made?
	#Relocation is one of the operational measures we are using to manage the #refugeecrisis

Table 17 (continued)

	'Open access is also a means of maximising the economic value and impact of publicly funded research.' GEOGHE-GAN-QUINN
	Supporting human rights-based migration management and asylum system in #Libya
	#Environment: Impact assessment of projects now simplified
	#Employment and Social Developments: growing divergence and higher risks of long-term #exclusion
	#SMEs: Better access to finance and boosting #entrepreneurship
Strategic Investments	8.1 billion investment in research and innovation to create growth and jobs
	8.1 billion investment in research and #innovation to create growth and jobs
	Horizon 2020: Commission proposes 80 billion investment in research and innovation, to boost growth and jobs
	We welcome @EUCouncil's approval of the Recovery and Resilience Plan for Bulgaria. This approval paves the way for disbursements of 6.3 billion to Bulgaria to foster its green transition and economic recovery from the COVID-19 pandemic
	European Fund for Strategic Investments is set to trigger investments of 251bn across Europe, after @EIB approved additional 1.8bn financing
	Did you know around 967,000 SMEs are set to benefit from #investEU funds? The Investment Plan is expected to trigger 424 billion in investments and is contributing to closing the investment gap caused by the economic crisis
	A new Economic Recovery Plan will mobilise up to 600 million in assistance, grants and investments for the Republic of Moldova. This will support the country's reforms and recovery, investing in the economy, connectivity, education and employability. #StrongerTogether
	We are supporting small and medium businesses in Luxembourg: 40 million in loans will help innovative firms turn their ideas into concrete projects. This support takes place under #investEU, the Investment Plan for Europe
	We have granted 56.3 million to support recovery projects in France and Germany under #ReactEU. The funds will support economic recovery, accelerate digital and green transitions and help young people affected by the pandemic.#NextGenerationEU
	We are supporting a 35 million @EIB loan to a social economy enterprise in Spain. It will allow the enterprise to create jobs for people with disabilities and to invest in improving energy efficiency
Health	We are funding the global fight against COVID-19. We have raised almost 16 billion with our partners under the Coronavirus #GlobalResponse. Team Europe has pledged 2.2 billion for COVAX, the worlds facility for universal access to vaccines.#EUDataCrunch

Table 17 (continued)

	We have secured a robust portfolio of 2.3 billion doses of the most promising #COVID19 vaccines using different technologies, more than enough to protect the entire EU population.#SafeVaccines #VaccinesWork
	Under the Coronavirus Global Response Initiative, President @vonderleyen rallied the international community to raise almost 16 billion to develop vaccines, diagnostics and treatments and make them available at affordable prices globally .#UnitedAgainstCoronavirus
	6 June Together with @GlblCtzn, we mobilised additional 6.15 billion to help develop and ensure access to COVID-19 vaccines and treatments around the world. The #vdLCommission introduced a new vaccine strategy to help develop safe and effective vaccines. #StrongerTogether
	We have signed a second contract with a pharmaceutical company to purchase vaccines for #coronavirus. The agreement with @sanofi-@gsk reflects our commitment to ensuring equitable access to vaccines for our citizens and the world's most vulnerable people. #StrongerTogether
	Vaccine against the #coronavirus is difficult to develop and requires money. With the #GlobalResponse we are joining forces with global partners to mobilise resources to develop a vaccine and deploy it in the entire world. Learn more about the event
	COVID-19 vaccines prepare your immune system to defend itself against the disease. There are several types of COVID-19 vaccines which work in different ways. #VaccinesWork! Learn more about how
	#TeamEurope has donated more than 350 million COVID-19 vaccine doses worldwide. And we will do more. Because we will only control this pandemic if we fight it in every corner of the world.#StrongerTogether #COVAX
	We do all we can to get #COVID19 vaccines to Europeans as quickly as possible. More than 43 million doses have been administered in Europe so far. 4 million additional doses of BioNTech-Pfizer vaccine will help tackle the virus hotspots across the EU. #SafeVaccines
	The world is coming together to defeat the coronavirus. We have so far raised 5.4 billion in initial funding for affordable and accessible diagnostics, treatments and vaccines.#UnitedAgainstCoronavirus #GlobalResponse #StrongerTogether
Citizen Rights and Integration	EU citizens are looking forward to travelling again this summer. The EU Digital COVID certificate allows them to do so in a free and safe way. We urge EU countries to get fully ready, so the system will be fully up and running on 1 July. #EUCOVIDcertificate
	Today EU-wide rules on electronic identification take effect. Europeans and businesses will be able to access online services even from a different EU country. It can also save businesses and administrations over 11 billion a year!
	Simpler,faster &safer Customs Union: New EU rules come into force today

Table 17 (continued)

	Today new rules to protect whistleblowers enter into force new EU-wide standards for people reporting breaches of EU law. From this day on, EU countries have 2 years to transpose rules into their national legislation. Learn more
	The.eu domain gives Europe its own Internet identity. Our updated rules will now allow all EU citizens residing outside the EU to register a.eu domain. More here
	Our #SingleMarket lets you do business across the EU. New rules adopted today will provide clear harmonised procedures for companies to merge, divide or move within the Single Market
	#OnThisDay 57 years ago the first regulation on free movement of workers in the entered into force. As an EU citizen, you have the right to work (and to look for a job) in any other EU country! Learn more
	#EUchat EU citizens have right to work in another EU country under same conditions as nationals. Is this right granted only to EU citizens?
	Familiar with your social rights? As a European working in the EU, you are entitled to a minimum of 4 weeks paid leave, social security protection and safety guarantees in your workplace. So, use your rights! #EUandME
	You can live and work in two different EU countries without breaking a sweat! The EU rules on social security make sure you can move around Europe and not lose out on your social security rights. Get all the details in the quick explainer
Human Rights	Today, we celebrate the 25th anniversary of the UN Convention on the rights of the child #UniversalChildrensDay
	Every human should be free to love who they want without fear of discrimination
	Every child should have the opportunity to aim for the stars. We work to prevent violence against children, promote child-friendly justice, defend their right to be heard, and much more. More
	International Day of Zero Tolerance against Female Genital Mutilation: Just one victim is too many! #zerofgm
	"All human beings are born free and equal in dignity and rights." Article 1 of the Universal Declaration of Human Rights #WeRemember
	On #ChildrensDay we stand united in our determination to protect and to promote the rights of all children everywhere. It is our collective task to do everything we can to make sure these rights are respected and ensured for every child. Statement
	Everyone has the right to the protection of personal data
	30 years of the UN Convention on the Rights of the Child. We celebrated the 30th anniversary of the UN Convention on the Rights of the Child. It's the most widely ratified human rights treaty in history that helped transform children's lives around the world for the better

Table 17 (continued)

	We are making sure everyone can use their right to free movement
	People with disabilities have equal rights and are entitled to full participation in society
Digital Policy	Digital Services Act Binding EU-wide obligations will apply to all digital services that connect consumers to goods, services, or content. The new framework will rebalance the rights and responsibilities of users, intermediary platforms, and public authorities.#DigitalEU
	You should stay in control of your data. We propose new rules on data governance to: Increase trust in data sharing Facilitate reuse of public sector data Give people more control Create neutral mediators. More
	Today we are proposing a set of new rules for all digital services: the Digital Services Act and the Digital Markets Act. We want to make sure users have access to a wide choice of safe products and services online, and that businesses compete fairly and freely. #DigitalEU
	We welcome agreement on our proposal to close important security gaps. The new rules will ensure that border guards and police officers have access to the right information when and where they need it
	The new Regulation on the free flow of non-personal data has entered into force. It will allow data to be stored and processed everywhere in the EU without unjustified restrictions. See new guidance to help users, especially SMEs
	Want to know more? Digital Services Act: Digital Markets Act: New online rules for users: New online rules for platforms: New online rules for businesses
	New #EU #foodlabelling rules—same #food info available in stores should be available online
	Personal data will travel safely between the EU and Japan. This is the world's largest area of safe data flows. EU companies will also benefit from privileged access to a market of 127 million consumers
	Europeans and businesses in favour of a fair data economy, according to results of the open public consultation on the Data Act. This Act will allow users to have more control over the data they generate, clarifying who can use it and for which purposes. #DigitalEU
	To keep people informed on health and travel measures, we launched Re-open EU, with accurate and up to date info on health measures and travel restrictions across the EU and in some partner countries
Digital and Green Transition	75% of total greenhouse gas emissions in the EU comes from the energy sector. We need to clean our energy systems to reach our climate goals, and we set two new binding targets by 2030: 40% of renewables in the EU energy mix 36% energy efficiency target#EUGreenDeal

Table 17 (continued)

	Think green. Think #EUGreenDeal. This month we presented new proposals on energy, transport and taxation to meet our 2030 climate targets. These will transform our economy and society, and create opportunities for innovation and jobs. Learn more
	Energy efficiency is key to meet our #EUGreenDeal goals.EU countries should reduce primary consumption by 39% by 2030. This will: cut our emissions ease the pressure on the environment reduce the need for energy to support our way of life decrease our energy bills
	Buildings are the single largest energy consumer in Europe. Our proposal on the energy performance of buildings aims to boost the rate of energy renovation across the EU. Stimulating renovation of homes and buildings also creates new job opportunities.#EUGreenDeal
	More 5 G and Gigabit networks across the EU. We invest more than 1 billion to improve Europes digital connectivity infrastructures under the Connecting Europe Facility. Secure and fast connectivity is the key on which we will build Europes Digital Decade.#DigitalEU
	Today we have adopted the Partnership Agreement for 2021–2027 with Sweden. It means the deployment of 2.2 billion in Sweden for: A circular and export-oriented green economy A smarter, digital and connected economy An inclusive labour market#EUFunds
	Our #EnergyUnion strategy aims to ensure that Europe's energy supply is safe, viable and accessible to all. Our new report shows progress made on bringing about a transition to a low-carbon and competitive EU economy. More
	The EU remains among the leaders in green high-value patents technology and green patents in energy intensive industries, according to the new EU Industrial R &D Investment Scoreboard. A reflection of our transformation towards climate neutrality.Check the interactive website
	Italy has proposed to invest in improving the energy efficiency of residential buildings. This will be a big boost for the green transition and it will also create jobs and growth!#MakeItReal #NextGenEU
	Open Sustainable Assertive This is what #EUtrade should be like in order to face the current challenges. Our new EU trade strategy aims at restoring growth and job creation while supporting the green and digital transition. More
Energy, Sustainability and Climate	The main features of the new #LIFEprogramme: clean energy nature and biodiversity circular economy and climate change mitigation simple and flexible Learn more
	The #GlobalGateway is our strategy to boost smart, clean and secure links in digital, energy and transport and also strengthen health, education and research systems across the world. Learn more#AfricaEU

Table 17 (continued)

	We have ambitious #ClimateAction objectives. Our strategy on #SustainableFinanceEU sets out a roadmap to allow investors and individual citizens to make positive choices for a greener and cleaner economy
	Sustainable batteries will contribute to the green transition and help us reach climate neutrality by: powering clean electric vehicles support our drive towards renewable energy reducing their carbon footprint #EUGreenDeal
	75% of total greenhouse gas emissions in the EU come from the energy sector. To meet our 2030 climate objectives and a climate-neutral continent by 2050, we must reshape the EUs energy system. Discover our new renewable energy targets with #EUDataCrunch
	Massive investments will be needed to transition to a climate-neutral economy. Today in Brussels we discuss #SustainableFinanceEU and our guidelines for climate-reporting for companies. Learn more and follow live
	A green and digital transition A new health programme A revitalised single market A reinforced rescEU A fair and inclusive recovery for all. With #NextGenerationEU, we want to invest in our future. More
	It is all about turning over a new leaf. Change We need change a transformational change. With the #EUGreenDeal, we aim to transform our economy and society for a sustainable, climate-neutral future by 2050, while at the same time leaving no one behind
	Europe needs to decarbonise the transport sector. We propose a transition to smart and sustainable mobility across Europe and in urban areas. Faster rail connections, improved public transport, and better infrastructure for walking and cycling.#EUGreenDeal
	Europe will become the first climate-neutral continent by 2050we will produce no more greenhouse gases than our ecosystems can naturally absorb. With #NextGenEU, we will invest in energy efficient buildings environmentally-friendly tech greener vehicles#MakeItReal
Identity, Culture and Citizen Engagement	It is time to decide what the European Union at 27 should be. Whatever happens, it must be a Europe built by Europeans. Take part in the online public consultation on the #FutureofEurope, addressed to all Europeans
	Today is the Dutch National Day: Fijne Koningsdag We stand with our friends in the Netherlands as the orange spirit stays indoors... for now. In these challenging times for all of Europe we are #StrongerTogether
	This week the first citizens' event of the Conference on the Future of Europe took place. The conversation on the future has started! Join in! Take a look at the topics to discuss: #TheFutureIsYours
	It is only together that we can build our Union of tomorrow. Today, we set out our ideas for shaping the Conference on the Future of Europe. This is our chance to show people that their voice counts in Europe. Our contribution

Table 17 (continued)

	Now is your turn to take a seat in shaping Europe's future!The third and final day of the Citizens' Panel in Florence is taking place as we speak. Submit your ideas here #TheFutureIsYours
	Future of Europe Debate: @VivianeRedingEU debates with citizens in #Sofia tomorrow morning! Join in with #eudeb8
	Your ideas, hopes and dreams are essential to the future. Join us live for the first plenary of the Conference on the Future of Europe!Tune in and learn how you can take part!#TheFutureIsYours
	Today is #EuropeDay! On this day in 1950, the Schuman Declaration proposed the creation of a European Coal and Steel Community, the European Unions ancestor. Only through unity and solidarity are we able to maintain the European project. Together we are stronger
	"It is only by remembering that we can build a brighter future together. Europe has itself been the scene of some of the worst atrocities known to man. Just as in Africa, time can never erase the darkest hours" @JunckerEU at #Kwibuka25 Full speech
	How do you see Europes future in the world? 200 Europeans, a third of which under 25, will meet this weekend in Strasbourg. They will debate the role of the EU in the world and how the EU should deal with migration. Join the debate. Submit your ideas now!#TheFutureIsYours
Citizen Initiatives	Today we celebrate European Citizens' Initiative Day! This year we focus on global trends in voting and participation. See how the European Citizens' Initiative creates a real sense of European identity and influence EU policies
	Today is European Day of Justice! A day to inform European citizens on their rights and to raise awareness of EU law
	Take a look at the ideas raised by EU citizens on the European citizens initiative
	1 million EU citizens can promote change in EU policies with the European Citizens' Initiative! #EUTakeTheInitiative
	To launch a European citizens initiative, it takes 7 EU citizens living in at least 7 different Member States who are old enough to vote. 1 million signatures will push the start button on EU policy-making. Take the Initiative!
	How many people got involved in EU law-making by using the European Citizens Initiative? #EUTakeTheInitiative
	Today is a good day for Europe! Ukraine and Moldova are granted the candidate status to become part of the European family. Georgia is given the perspective to become a member of the EU.#EUCO
	Do you know how many people it takes to launch a Europeans citizens' initiative? 7 EU citizens living in at least 7 different EU countries who are old enough to vote. Learn more and get involved

Table 17 (continued)

	#DiscoverEU is an initiative of the European Union that gives 18-year-olds the opportunity to discover Europe through learning experiences. This year 20,000 more young people will participate under latest round, which attracted some 75,000 applications from across Europe. More
	Are you planning a European citizens initiative, or already promoting one?Join the European Citizens Initiative Forum and learn how to: Prepare and register your initiative Successfully collect 1 million signatures Find partners across the EU#EUTakeTheInitiative
Visions for the Future	"The future will be what we make it. And Europe will be what we want it to be. Let's make it strong. And let's build the world we want to live in."—President @vonderleyen in her #SOTEU address
	Shaping a greener Europe with you. Today, we are launching the European Climate Pact to find climate solutions together, because when it comes to tackling climate change, anyone can take action for our planet.#EUGreenDeal
	Panel lists of the last citizens' panel will meet this week in Dublin to finalise their proposals for Europe's future on: Education, culture, youth, sport Digital transformation A stronger economy, social justice, jobs More: #TheFutureIsYours
	"The future will be what we make it. And Europe will be what we want it to be. Let's make it strong. And let's build the world we want to live in."—President @vonderleyen in her #SOTEU address
	Se sembra impossibile, allora si pu fare. This was the spirit of Europes founders and this is the spirit of Europes next generation. This is the future of Europe. Lets make it stronger together. Viva lEuropa.#SOTEU
	Only by finding strength in unity, will Europe be able to influence the course of world events. This week, we made recommendations for the EUs next strategic agenda 2019–2024. Read more on #EURoad2Sibiu
	We want a digital transformation that works for all. 6 months ago, we announced our strategy to shape Europes digital future. Next week during her first State of the Union address, President @vonderleyen will present the priorities for the year ahead. #DigitalEU #SOTEU
	Worried about what the future holds for the environment, the economy or even democracy in Europe?People have already started discussing these and other topics in the Conference on the Future of Europe. Join the discussion! #TheFutureIsYours
	Some Europeans say we should have a green education for all generations. What do you think?Join the citizen-led debate on Europe's future. Submit your ideas now!#TheFutureIsYours

Table 17 (continued)

	Moving towards a cleaner and healthier future for all. Join the Web Summit session with Executive Vice-President @TimmermansEU to know more about our plans for a green future for Europe. #EUatWebSummit #EUGreenDeal
Internal Governance	EU-NATO cooperation joint declaration: We reaffirm our joint commitment to one another. President @JunckerEU, @eucopresident Tusk and NATO Secretary General @jensstoltenberg
	Press Statement of President Charles Michel of the European Council and President Ursula von der Leyen of the European Commission on Russia's unprecedented and unprovoked military aggression of Ukraine
	#EUCO Statement by the Presidents of the European Council, the Commission and the Euro group on #Greece
	European Parliament has elected Jean-Claude Juncker the next President of @EU_Commission. Follow @JunckerEU #EC2014
	Speech by President @EU_Barroso at the EP plenary debate on the Irish Presidency of the Council @eu2013ie @Europarl_EN
	Tomorrow, the EU and NATO will reaffirm their joint commitment to one another, with a joint declaration on EU-NATO cooperation, signed by President @JunckerEU, @eucopresident Tusk and NATO Secretary General @jensstoltenberg
	President @JunckerEU addresses the #EPlenary session of the European Parliament on the Commission's preparations ahead of the #EUCO of October
	Press conference by @MichelBarnier the European Commission's Chief Negotiator, and David Davis, UK Secretary of State for Exiting the European Union #Brexit #Article50
	#EU #Commission Pdt #Barrosoexpects clear statement of Friday European Council on #Egypt, #Tunisia, #Belarus
	Statement by the EU leaders following the special meeting of the European Council on #migrationEU
Trade, Partnerships, and Law	.@EU_Commission calls on the U.S. to restore trust in EU-US data flows
	.@EU_Commission requests Member States to comply with EU law when regulating #gambling services
	State aid | Hungary: @EU_Commission finds food chain inspection fee and tax on tobacco sales in breach of EU rules
	#Stateaid: @EU_Commission clarifies scope of EU State aid rules to facilitate public investment
	.@EU_Commission adopts partnership agreement w/ France on EU Structural and Investment Funds for growth and jobs
	Working conditions: @EU_Commission reviews application of EU rules on temporary agency work
	.@EU_Commission calls on MS to implement rules for people sentenced in other EU country

Table 17 (continued)

	Tackling #Tax Avoidance: @EU_Commission tightens key EU corporate tax rules
	EU's top 2014 goods trade partners: USA for exports, China for imports via @EU_Eurostat
	Information and consultation at work: @EU_Commission evaluates #EU rules
Solidarity and Emergency Response	As humanitarian needs continue to rise, the EU is providing direct operational support to aid organisations working inside Ukraine. The EU has set up two warehouses in Ukraine, and we are contracting trucking services to get humanitarian aid to those in need.#StandWithUkraine
	The EU continues to support people in Haiti who are suffering the consequences of the devastating earthquake that hit the country. An EU Humanitarian Air Bridge operation is delivering more than 125 tonnes of life-saving materials
	The EU continues to stand in solidarity with the Palestinian people. Over 2 million Palestinians in the Occupied Territories and Gaza Strip are in need of humanitarian assistance. The EU will support those vulnerable people with 25 million in humanitarian aid.#EUSolidarity
	The EU Emergency Response Coordination Centre continues to support the repatriation of EU citizens stranded abroad during the #coronavirus outbreak. More than half a million people have been flown back home. More
	Following the tragic blast in Beirut over 2 months ago, we have supported the response with: 54 tonnes of supplies via 3 EU humanitarian air bridge flights64 million in emergency funding+1,000 tonnes of in-kind aid by EU countries facilitated by EU civil protection
	The EU logistics hub in Moldova helps address the immediate needs of people fleeing war.EU humanitarian partners can use the warehouse to stock and deploy supplies to support civilians fleeing the war in Ukraine.#EUsolidarity #StandWithUkraine
	The EU is allocating a further 50 million to support Ukrainians fleeing war as their humanitarian needs remain extremely high. This new funding will provide: emergency healthcare safe drinking water shelter cash assistance support against gender-based violence
	Civilians in eastern Ukraine are still caught up in a conflict that started in 2014. The EU announced additional 10 million humanitarian funding to help the most vulnerable people living close to the front line prepare for winter. Read more
	When a disaster strikes, the EU is ready to lend a helping hand. We propose 86.7 million to provide relief to people hit by natural disasters in France and Greece in 2020. More
	The EU is stepping up its long-standing humanitarian assistance to those most in need. We provide 39 million in humanitarian aid to support displaced and conflict-affected people in Bangladesh and Myanmar. More
Financial Aid	5 million in humanitarian aid for the Burundian crisis

Table 17 (continued)

	29 million in humanitarian aid for the most vulnerable populations in #Somalia
	#Sudan: 106 million package to directly assist some 4.8 million people needing urgent assistance amid displacement
	Additional 20 million to support Jordan in #Syria refugee crisis
	62 million in humanitarian aid to support Syrians displaced inside the country
	#MigrationEU—An additional 9.3 million in emergency funding to #UNHCR to provide support to refugees in Greece
	We have awarded an additional 3 million in emergency assistance to support Spain in responding to the recent migratory challenges
	#Turkey refugees | 714 million in humanitarian aid and 50 million for skills training. Our largest aid operation
	16.2 million of EU Solidarity Funds allocated to #Greece and #Bulgaria in the wake of natural disasters
	#migrationEU: additional 5 million for emergency humanitarian assistance in #Calais #France
Press conferences and Statements	Statement of President #Barroso after the meeting with Prime Minister Netanyahu
	LIVE: Watch the press conference with Prime Minister Yatseniuk of #Ukraine
	Watch live: @EU_Commission President @JunckerEU in the press room now
	Statement by President @ BarrosoEU following his meeting with Mr#Fico, Prime Minister of #Slovakia
	12h50 (TBC): Joint press conference by Commission President #Barroso and Hungarian PM Viktor #Orban in Budapest
	11h30 CET: Press conference with #Barroso, @JerzyBuzek and #Polish PM #Tusk on Polish presidency, watch live
	Statement by Cssr tefan #Fle following his meeting with Mr Habib Essid, Minister of Interior of #Tunisia
	Statement by Vice President #Rehn ahead of the# G20 Finance Ministers and Central Bank Governors meeting in #Moscow
	Live now: Press conference following #EUCO with President @JunckerEU, @eucopresident Tusk and Estonian Prime Minister Ratas
	EU-Canada Summit: Press Conference LIVE
Charts, Links, Infographics	#teamJuncker activities for the week ahead (2–8 November 2015)
	More information
	More information here
	Have you voted yet? Check out the results on
	Take a look at the #TeamJunckerEU agenda for the week

Table 17 (continued)

	Check out the calendar with #TeamJunckerEU activities for the week
	Check out the calendar with #TeamJunckerEU activities for the week
	What will #TeamJunckerEU be doing this week? Check out the agenda
	Check out the #TeamJunckerEU calendar for the week
	What will #TeamJunckerEU be doing this week? Check out the agenda
Live Events	Watch with live interpretation here
	Watch with live interpretation here
	Watch with live interpretation here
	Press conference on 8th round #TTIP negotiations LIVE from 15:30 CET
	Find the schedule for today's live transmissions and press conferences here
	Press conference on #Ukraine now live
	Press conference with Vice-President @VDombrovskis LIVE from ± 12:15 (CET)
	Press point with @JunckerEU and @Poroshenko live NOW
	FOLLOW THE @JunckerEU PRESS CONFERENCE HERE
	Watch live now: signature ceremony and joint press briefing w/ @BarrosoEU and @Yatsenyuk_AP

Data availability The code used to perform the analyses is available at https://github.com/rbroc/eucomm-twitter. Terms and conditions for the Twitter Academic API do not allow sharing of raw Twitter data. We share extracted features in the GitHub repository.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Authors' Google Scholar Profiles RR: https://scholar.google.com/citations?user=EjB67tAAAAAJ. KL: Not available. MT: https://scholar.google.com/citations?user=hj8KVtgAAAAJ. LC: Not available.

References

1. Grande, E., & Hutter, S. (2016). In S. Hutter, E. Grande, & Kriesi, H. (eds.) *Introduction: European Integration and the Challenge of Politicisation* (pp. 3–31). Cambridge University Press. https://doi.org/10.1017/CBO9781316422991.002

2. Hooghe, L., & Marks, G. (2009). A postfunctionalist theory of European integration: From permissive consensus to constraining dissensus. *British Journal of Political Science, 39*(1), 1–23. https://doi.org/10.1017/S0007123408000409

3. Rauh, C. (2019). EU politicization and policy initiatives of the European Commission: The case of consumer policy. *Journal of European Public Policy, 26*(3), 344–365. https://doi.org/10.1080/13501763.2018.1453528

4. Barisione, M., & Michailidou, A. (2017). *Social media and European politics: Rethinking power and legitimacy in the digital era. Palgrave Studies in European Political Sociology*. Palgrave Macmillan.

5. Özdemir, S., & Rauh, C. (2022). A bird's eye view: Supranational EU actors on Twitter. *Politics and Governance, 10*(1), 133–145. https://doi.org/10.17645/pag.v10i1.4686

6. Zaiotti, R. (2020). The (UN) making of international organisations' digital reputation: The European Union, the "refugee crisis," and social media. In *Digital diplomacy and international organisations*. Routledge.

7. Usher, J. (1994). *The commission and the law. The European Commission* (pp. 146–68). Longman.

8. Bianchi, F., Terragni, S., Hovy, D., Nozza, D., & Fersini, E. (2021). Cross-lingual contextualized topic models with zero-shot learning. In *Proceedings of the 16th conference of the European chapter of the Association for Computational Linguistics: Main volume*, pp. 1676–1683. Association for Computational Linguistics, Online. https://doi.org/10.18653/v1/2021.eacl-main.143. https://aclanthology.org/2021.eacl-main.143. Accessed 2023-05-02.

9. Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2020). DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter. https://doi.org/10.48550/arXiv.1910.01108. arXiv:1910.01108. Accessed 2023-05-02.

10. Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., Drame, M., Lhoest, Q., & Rush, A. (2020). Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 conference on empirical methods in natural language processing: System demonstrations* (pp. 38–45). Association for Computational Linguistics, Online. https://doi.org/10.18653/v1/2020.emnlp-demos.6. https://aclanthology.org/2020.emnlp-demos.6. Accessed 2023-05-02.

11. Benoit, K., Munger, K., & Spirling, A. (2019). Measuring and explaining political sophistication through textual complexity. *American Journal of Political Science, 63*(2), 491–508. https://doi.org/10.1111/ajps.12423

12. Rauh, C. (2022). Clear messages to the European public? The language of European Commission press releases 1985–2020. *Journal of European Integration*. https://doi.org/10.1080/07036337.2022.2134860

13. Niţoiu, C. (2013). The narrative construction of the European Union in external relations. *Perspectives on European Politics and Society, 14*(2), 240–255. https://doi.org/10.1080/15705854.2013.785264

14. Manor, I. (2019). *The Digitalization of Public Diplomacy*. Paris: Springer. https://doi.org/10.1007/978-3-030-04405-3

15. Cassidy, J. (2018). Digital diplomatic crisis communication: Reconceptualising diplomatic signalling in an age of real time governance. Technical report, DigDiploROx Working Paper No. 3.

16. Hedling, E. (2020). Storytelling in EU public diplomacy: Reputation management and recognition of success. *Place Branding and Public Diplomacy, 16*(2), 143–152. https://doi.org/10.1057/s41254-019-00138-2

17. Moral, P. (2023). Restoring reputation through digital diplomacy: The European union's strategic narratives on twitter during the covid-19 pandemic. *Communication and Society, 36*, 241–269. https://doi.org/10.15581/003.36.2.241-269

18. Wright, K. A., & Guerrina, R. (2020). Imagining the European Union: Gender and digital diplomacy in European external relations. *Political Studies Review, 18*(3), 393–409. https://doi.org/10.1177/1478929919893935

19. Bianchi, F., Terragni, S., & Hovy, D. (2021). Pre-training is a hot topic: Contextualized document embeddings improve topic coherence. In *Proceedings of the 59th annual meeting of the Association for Computational Linguistics and the 11th international joint conference on natural language processing (Volume 2: Short Papers)* (pp. 759–766). Association for Computational Linguistics, Online. https://doi.org/10.18653/v1/2021.acl-short.96. https://aclanthology.org/2021.acl-short.96 Accessed 2023-05-02.

20. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems* (Vol. 30). Curran Associates, Inc.

21. Ferrara, E., & Yang, Z. (2015). Quantifying the effect of sentiment on information diffusion in social media. *PeerJ Computer Science, 1*, 26. https://doi.org/10.7717/peerj-cs.26

22. Jiménez-Zafra, S. M., Sáez-Castillo, A. J., Conde-Sánchez, A., & Martín-Valdivia, M. T. (2021). How do sentiments affect virality on Twitter? *Royal Society Open Science, 8*(4), 201756–201756. https://doi.org/10.1098/rsos.201756

23. Stieglitz, S., & Dang-Xuan, L. (2013). Emotions and information diffusion in social media-sentiment of microblogs and sharing behavior. *Journal of Management Information Systems, 29*(4), 217–248. https://doi.org/10.2753/MIS0742-1222290408

24. Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*(4), 977–990. https://doi.org/10.3758/BRM.41.4.977

25. Anderson, J. (1983). Lix and Rix: Variations on a little-known readability index. *Journal of Reading, 26*(6), 490–496.

26. Coleman, M., & Liau, T. L. (1975). A computer readability formula designed for machine scoring. *Journal of Applied Psychology, 60*, 283–284. https://doi.org/10.1037/h0076540

27. Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology, 32*, 221–233. https://doi.org/10.1037/h0057532

28. Gunning, R. (1969). The Fog Index after twenty years. *Journal of Business Communication, 6*(2), 3–13. https://doi.org/10.1177/002194366900600202

29. McLaughlin, G. H. (1969). SMOG grading—A new readability formula. *Journal of Reading, 12*(8), 639–646.

30. Kincaid, J. P., Fishburne, R. P. J., Rogers, R. L., & Chissom, B. S. Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel. Technical report. Institute for Simulation and Training (1975). https://apps.dtic.mil/sti/citations/ADA006655 Accessed 2023-05-03.

31. Smith, E. A., & Senter, R. J. Automated Readability Index. Technical report (1967). https://apps.dtic.mil/sti/citations/AD0667273. Accessed 2023-05-03.

32. Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods, 46*(3), 904–911. https://doi.org/10.3758/s13428-013-0403-5

33. Hansen, L., Olsen, L. R., & Enevoldsen, K. TextDescriptives: A Python package for calculating a large variety of metrics from text (2023). https://doi.org/10.48550/arXiv.2301.02057. arXiv:2301.02057 [cs]. Accessed 2023-05-02.

34. McNamara, Q., De La Vega, A., & Yarkoni, T. Developing a comprehensive framework for multimodal feature extraction. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining. KDD '17* (pp. 1567–1574). Association for Computing Machinery (2017). https://doi.org/10.1145/3097983.3098075. https://dl.acm.org/doi/10.1145/3097983.3098075. Accessed 2023-05-02.

35. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. RoBERTa: A robustly optimized BERT pretraining approach (2019). arXiv:1907.11692 [cs]. https://doi.org/10.48550/arXiv.1907.11692. Accessed 2023-05-02.

36. Barbieri, F., Camacho-Collados, J., Neves, L., & Espinosa-Anke, L. TweetEval: Unified benchmark and comparative evaluation for Tweet classification (2020). https://doi.org/10.48550/arXiv.2010.12421. arXiv:2010.12421 [cs]. Accessed 2023-05-02.

37. Huber, P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics, 35*(1), 73–101. https://doi.org/10.1214/aoms/1177703732

38. Chen, T., & Guestrin, C. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. KDD '16* (pp. 785–794). Association for Computing Machinery (2016). https://doi.org/10.1145/2939672.2939785. https://dl.acm.org/doi/10.1145/2939672.2939785. Accessed 2023-05-02.

39. Rocca, R., & Yarkoni, T. (2021). Putting psychology to the test: Rethinking model evaluation through benchmarking and prediction. *Advances in Methods and Practices in Psychological Science, 4*(3), 25152459211026864. https://doi.org/10.1177/25152459211026864

40. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems.* (Vol. 30). Curran Associates Inc.

41. Reimers, N., & Gurevych, I. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)* (pp. 3982–3992). Association for Computational Linguistics (2019). https://doi.org/10.18653/v1/D19-1410. https://aclanthology.org/D19-1410. Accessed 2023-05-02.

42. Röder, M., Both, A., & Hinneburg, A. Exploring the space of topic coherence measures. In *Proceedings of the eighth ACM international conference on web search and data mining. WSDM '15* (pp. 399–408). Association for Computing Machinery (2015). https://doi.org/10.1145/2684822.2685324. https://dl.acm.org/doi/10.1145/2684822.2685324. Accessed 2023-05-02.

43. Hoyle, A., Goel, P., Sarkar, R., & Resnik, P. Are neural topic models broken? (2022). https://doi.org/10.48550/arXiv.2210.16162. arXiv:2210.16162. Accessed 2023-05-02.

44. Zhou, H., Yang, Y., & Qian, W. Tweedie gradient boosting for extremely unbalanced zero-inflated data (2019). https://doi.org/10.48550/arXiv.1811.10192. arXiv:1811.10192. Accessed 2023-05-02.

Authors and Affiliations

Roberta Rocca[1] [ORCID] **· Katharina Lawall[2] · Manos Tsakiris[2,3] · Laura Cram[4]**

✉ Roberta Rocca
roberta.rocca@cas.au.dk

Katharina Lawall
katharina.lawall@rhul.ac.uk

Manos Tsakiris
manos.tsakiris@rhul.ac.uk

Laura Cram
laura.cram@ed.ac.uk

[1] Interacting Minds Centre, Department of Culture, Cognition, and Computation, Aarhus University, Aarhus, Denmark

[2] Centre for the Politics of Feelings, School of Advanced Study, University of London, London, UK

[3] Department of Psychology, Royal Holloway University of London, London, UK

[4] NRlabs Neuropolitics Research, School of Social and Political Science, University of Edinburgh, Edinburgh, UK