EPFL

Thèse n° 9138

How Words Move Hearts: Interpretable Machine Learning Models of Bias, Engagement, and Influence in Socio-Political Systems

Présentée le 17 octobre 2023

Faculté informatique et communications
Laboratoire de la dynamique de l'information et des réseaux 1
Programme doctoral en informatique et communications

pour l'obtention du grade de Docteur ès Sciences

par

Aswin SURESH

Acceptée sur proposition du jury

Prof. A. Argyraki, présidente du jury
Prof. M. Grossglauser, directeur de thèse
Prof. A. Chaintreau, rapporteur
Dr A. Carneiro Viana, rapporteuse
Prof. K. Aberer, rapporteur

École polytechnique fédérale de Lausanne

2023

"Words - so innocent and powerless as they are, as standing in a dictionary, how potent for good and evil they become in the hands of one who knows how to combine them."

— Nathaniel Hawthorne

To Appa, Amma, Kichu, and Harithakutty...

Acknowledgements

I am deeply indebted to many for the successful completion of this thesis. First and foremost, I thank my advisor, Professor Matthias Grossglauser, who accepted me and guided me through these years. He is one of the kindest and most brilliant people I know. He gave me almost total freedom in choosing my research topics, and his creative insights were a great source of inspiration in designing novel experiments. The doctoral journey can be quite intense emotionally. Whenever the going got tough, Matthias was always there to support me with words of reassurance and encouragement drawn from his years of experience. Thank you, Matthias, for being the best advisor anyone could ever hope for.

It was an honor to have Professor Augustin Chaintreau, Dr.Aline Carneiro Viana, Professor Karl Aberer, and Professor Katerina Argyraki as members of my thesis jury. Their valuable feedback helped me gain new perspectives on my current work and gave me ideas for future extensions. I sincerely thank all of them for their time and for their appreciation for our research.

Although he was not officially my advisor, Professor Patrick Thiran, who leads the sister lab within the INDY group, was almost a de facto co-advisor to me. During social events at the lab, I learned a lot from the discussions that we had regarding research and careers. We collaborated on a project related to law-making in the European Parliament, and he gave me helpful advice for the study on lobby influence. He also took an active part in the INDY group meetings and provided detailed feedback on the student projects that were contributions to the chapters of this thesis. Thank you for your guidance, Patrick.

All academic research is enabled by behind-the-scenes hard work of many non-academics and this thesis is no exception. Our lab secretaries, Patricia Hjelt and Angela Devenoge, patiently helped me with several administrative tasks, including hiring numerous student assistants and processing research and travel expenses. Our English editor, Holly Cogliati-Bauereis, ensured that the language in the chapters of this thesis (and the papers that they are based on) is correct, unambiguous, and conformant to accepted style, a complex process that we fondly refer to as 'hollification'. Our system administrators, Marc-André Lüthi and Yves Lopez, helped maintain the IT infrastructure of the lab, including the servers and clusters that were critical for running our experiments. This thesis would

not have been possible without their help, and I thank them all.
I thank Dr. Victor Kristof who was a senior colleague at the lab and a mentor to me. I drew deep inspiration from his work on socio-political systems, climate change, and discrete-choice models, for my own research in these topics. We collaborated on the projects related to the European Parliament discussed in Chapters 4 and 5 of this thesis. I also learned a lot from him about supervising students, writing, and presenting technical topics, in an effective and engaging manner. Thank you, Victor, for being a great mentor and friend. I also thank Dr. William Trouleau, another senior colleague and mentor. I had the opportunity to work with him as a teaching assistant for the Internet Analytics course at EPFL. His diligence and attention to detail were truly inspiring.
The work presented in this thesis relies heavily on the contributions of many great students that I had the opportunity to supervise. In particular, I thank Chi-Hsuan Wu, Lazar Milikic, Francis Murray, Yurui Zhu, Lazar Radojevic, Francesco Salvi, and Antoine Magron, who are also co-authors of the publications that the chapters are based on. They and many others did practical work including scraping and cleaning datasets, implementing models, running experiments, and generating plots; this work was essential, although perhaps not the most exciting component of research. Thank you all for your help. I hope that I was able to contribute to your training through my supervision.
I thank all my colleagues at INDY, as they were a great source of support and encouragement throughout this doctoral journey. Thank you Saber, Surya, Maximilien, Alexandre, Lucas, Farnood, Daniyar, Greg, Arnout, Lars, Mladen, Sadegh, Saeed, Paula, and Daichi. In particular, I thank Mahsa with whom I shared an office all these years and whose PhD was contemporaneous with mine. I will cherish our many conversations about the PhD experience and life in general; they were a welcome break from work.
I thank my friends outside the lab, especially the Indian student community at EPFL. They made me feel at home while being thousands of kilometers away. A special thank you goes to Sachin whose PhD was contemporaneous with mine, and who was the only one to whom I could speak in my mother tongue (Malayalam). We shared an apartment during this time and supported each other through our respective journeys.
Finally, I am deeply grateful to my family. Life as a PhD student can be quite stressful at times. The unconditional love and support that I received from my parents Suresh and Premaletha, my brother Athul, and my wife Haritha, helped me tide over such episodes and successfully complete this thesis. Thank you for always standing by me.

Lausanne, 28 September 2023 A. S.

Abstract

We study socio-political systems in representative democracies. Motivated by problems that affect the proper functioning of the system, we build computational methods to answer research questions regarding the phenomena occurring in them. For each phenomenon, we curate novel datasets and propose interpretable models that build upon prior work on distributional representations of text, topic models, and discrete choice models. These models provide valuable insights and enable the construction of tools that could help solve some of the problems affecting the systems.

First, we look at the problem of subjective bias in documents on the Web and in media. We curate a novel dataset based on Wikipedia's revision history that contains pairs of versions of the same Wikipedia article, where the subjective bias in one version has been corrected to generate the other version. We train a Bradley-Terry model that uses text features to perform a pairwise comparison of bias between these versions. We show that we can interpret the parameters of the model to discover the words most indicative of bias. Our model also learns to compute a real-valued bias score for documents. We show that this score can be used as a measure of bias across topics and domains not seen in training, including in the media, political speeches, law amendments, and tweets.

Second, we infer effective strategies for improving user engagement in social media campaigns, taking the example of tweets about climate change. We build an interpretable model to rank tweets on the basis of predicted engagement by using their topic and metadata features. The ranking framework enables us to avoid the influence of confounding factors such as author popularity. We make several recommendations for the optimization of engagement, based on the learned model parameters, such as talking about mitigation and adaptation strategies, instead of projections and effects.

Third, we study the influence of interest groups (lobbies) on parliaments, taking the European Parliament (EP) as an example. We curate novel datasets of the position papers of the lobbies and speeches of the members of the EP (MEPs), and we match them to discover interpretable links between lobbies and MEPs. In the absence of ground-truth data, we indirectly validate the discovered links by comparing them with a dataset, which we curate, of retweet links between lobbies and MEPs and with the publicly disclosed meetings of MEPs. An aggregate analysis of the discovered links reveals patterns that

follow ideology (e.g., the center-left political group is more associated with social causes). Finally, we study the law-making process within the EP. We mine a rich dataset of edits to law proposals and develop models that predict their acceptance by parliamentary committees. Our models use textual and metadata features of the edit, and latent features to capture the interaction between parliamentarians and laws. We show that the model accurately predicts the acceptance of the edits. Furthermore, we interpret the parameters of the learned model to derive interesting insights into the legislative process. We show that, among other observations, edits that narrow the scope of the law and insert recommendations for actions are more likely to be accepted than those that insert obligations.

Key words: computational social science, natural language processing, discrete choice models, social systems, bias, law-making, Wikipedia, Twitter, word embeddings, topic models

Résumé

Nous étudions les systèmes sociopolitiques dans les démocraties représentatives et construisons des méthodes informatiques pour répondre aux questions de recherche concernant les phénomènes qui s'y produisent, motivés par des problèmes qui affectent le bon fonctionnement des systèmes. Pour chaque phénomène, nous rassemblons de nouveaux jeux de données et proposons des modèles interprétables qui s'appuient sur des travaux antérieurs concernant les représentations distributionnelles de textes, les modèles de sujets et les modèles de choix discrets. Ces modèles fournissent des informations précieuses et permettent la construction d'outils qui pourraient aider à résoudre certains des problèmes affectant les systèmes.

Tout d'abord, nous examinons le problème de la partialité subjective dans les documents sur le web et dans les médias. Nous créons un nouveau jeu de données basé sur l'historique des révisions de Wikipédia contenant des paires de versions du même article de Wikipédia où le biais subjectif d'une version a été corrigé pour générer l'autre version. Nous entraînons un modèle Bradley-Terry utilisant des caractéristiques textuelles pour effectuer une comparaison par paire des biais entre ces versions. Nous montrons que nous pouvons interpréter les paramètres du modèle pour découvrir les mots les plus révélateurs de la partialité. Notre modèle apprend également à calculer un score de partialité pour les documents. Nous montrons que ce score peut être utilisé comme mesure de la partialité dans des sujets et des domaines qui n'ont pas été vus lors de l'entraînement du model, y compris les médias, les discours politiques, les amendements à la loi et les tweets.

Deuxièmement, nous déduisons des stratégies efficaces pour améliorer l'engagement des utilisateurs dans les campagnes de médias sociaux, en prenant l'exemple des tweets sur le changement climatique. Nous construisons un modèle interprétable pour classer les tweets sur la base de l'engagement prédit en utilisant leur sujet et les caractéristiques des métadonnées. Le système de classement nous permet d'éviter l'influence de facteurs confondants tels que la popularité de l'auteur. Nous formulons plusieurs recommandations pour optimiser l'engagement sur la base des paramètres du modèle appris, comme le fait de parler de stratégies d'atténuation et d'adaptation plutôt que de projections et d'impacts.

Troisièmement, nous étudions l'influence des groupes d'intérêt (lobbies) sur les parlements,

en prenant l'exemple du Parlement européen (PE). Nous rassemblons des nouveaux jeux de données de documents sur les positions des lobbies et sur les discours des membres du Parlement européen (MPE) et les mettons en correspondance pour découvrir des liens interprétables entre les lobbies et les MPE. En l'absence de données de référence, nous validons indirectement les liens découverts en les comparant à un ensemble de données que nous conservons sur les liens de retweet entre les lobbies et les députés européens, ainsi qu'aux réunions des députés européens divulguées publiquement. Une analyse globale des liens découverts révèle des schémas qui suivent l'idéologie (par exemple : le groupe politique de centre-gauche est associé aux lobbies pour les causes sociales).
Enfin, nous étudions le processus législatif au sein du Parlement européen. Nous exploitons un riche ensemble de données sur les modifications apportées aux propositions de loi et développons des modèles qui prédisent l'acceptation d'une modification par les commissions parlementaires. Nos modèles utilisent des caractéristiques textuelles et des métadonnées de l'édition et des caractéristiques latentes qui capturent l'interaction entre les parlementaires et les lois. Nous montrons que le modèle prédit avec précision l'acceptation des modifications. En outre, nous interprétons les paramètres du modèle appris pour en tirer des informations intéressantes sur le processus législatif. Entre autres observations, nous montrons que les modifications qui réduisent le champ d'application de la loi et insèrent des recommandations d'action ont plus de chances d'être acceptées que celles qui insèrent des obligations.

Mots clés : sciences sociales computationnelles, traitement automatique du langage naturel, modèles de choix discrets, systèmes sociaux, partialité, législation, Wikipédia, Twitter, plongement lexical, modèles thématiques

Contents

Mathematical Notation

Symbol	Description
x	Plain lowercase letters denote scalar values.
$\mathbf{x} = [x_i]$	Boldface lowercase letters denote column vectors.
$\mathbf{X} = [x_{ij}]$	Boldface uppercase letters denote matrices.
$\mathcal{X}$	Calligraphic uppercase letters denote sets.
$\mathbb{R}, \mathbb{R}_{>0}, \mathbb{N}$	Number types: real, positive real and natural numbers, respectively.
$i \succ j$	Pairwise comparison outcome "i is chosen over j".
$i \succ \mathcal{A}$	Multiway comparison outcome "i is chosen over all items in $\mathcal{A}$" ($i \notin \mathcal{A}$) .
$i \succ \mathcal{A} - \{i\}$	Multiway comparison outcome "i is chosen among items in $\mathcal{A}$".
$P(\mathcal{A})$	Probability of the event $\mathcal{A}$.
$\mathbb{1}_{\mathcal{A}}$	Indicator variable of the event $\mathcal{A}$.
$\exp(x)$	Exponential function $\exp(x) = e^x$.
$\sigma(x)$	Sigmoid function $\sigma : \mathbb{R} \to [0, 1]$, $\sigma(x) = 1/[1 + \exp(-x)]$.

1 Introduction

1.1 Motivation and Background

We humans are social animals. Modern human civilization is organized around social systems where large numbers of people collaborate to achieve common goals. In this thesis, we focus on arguably the largest and most important such system, namely the *socio-political system* in *representative democracies*. In this section, we briefly describe the structure of these systems and the phenomena that occur in them, we identify problems that affect their proper functioning and state our research questions inspired by these problems that we seek to answer in this thesis.

We give a highly simplified and idealized picture of such systems in Figure 1.1, showing the main entities and the information flows between them. *Citizens* of representative democracies receive information from reputed *media* outlets and some independent sources on the *Web*, such as Wikipedia, which they trust to be relatively objective (as opposed to information from state-controlled media that is often seen as biased in favor of the government) (Elmimouni et al., 2022; Pew Research, 2011). They use this information to guide their voting decisions during the elections (Dewenter et al., 2019; Stanford Internet Observatory, 2021) that determine the composition of the *Parliament*, the body that is responsible for making laws. New laws are proposed by the *Executive* to the Parliament that has the power to make amendments to them before finally deciding to accept (pass) or reject them. The laws passed by the Parliament are then enforced by the Executive.

In addition to this Citizen-Parliament-Executive link that is often explicitly set out in countries' constitutions, *industry associations* that represent the interests of groups of companies working in the same industry (e.g. cement manufacturers, steel makers, refineries, etc.) often try to influence the law-making process to protect and advance their interests. They do so through the Executive before an official law proposal is made, and through the Parliament afterward (Rasmussen, 2015). Other groups such as *non-governmental organizations* (NGOs) also try to exert their influence, usually

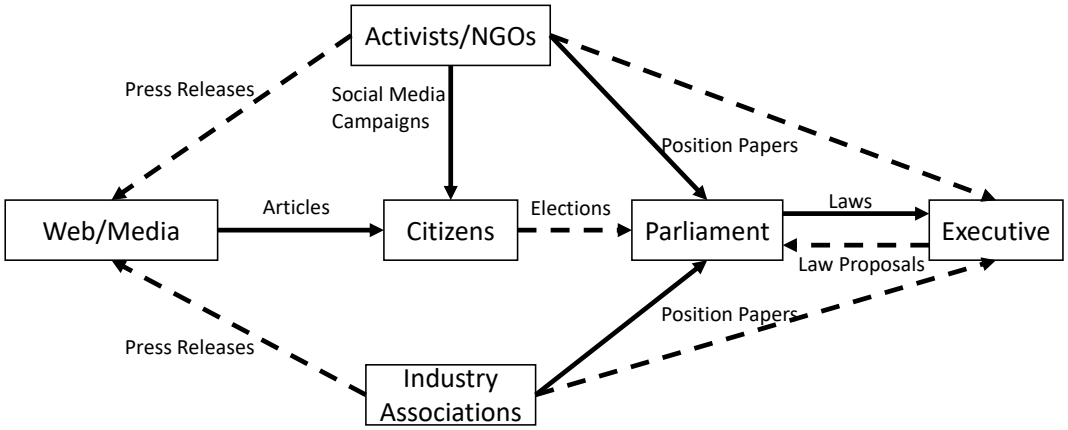

Figure 1.1: The socio-political system in representative democracies. This is a highly simplified picture; several entities and information flows are omitted for clarity. The flows we consider in this thesis are shown by solid arrows.

for socially relevant causes such as protecting human rights or the environment. We collectively refer to industry associations and NGOs as interest groups or lobbies, as they are more commonly known.

Traditionally, lobbying activity has been largely opaque to ordinary citizens, although recent initiatives, such as the EU Transparency Register (European Union, 2011), have brought some visibility to the process. In addition to covert forms of lobbying involving closed-door meetings, many lobbies publish their views on various policy questions in the form of position papers on their website and issue press releases to the media (Dialer & Richter, 2019). Besides influencing parliamentarians directly, NGOs and individual activists also conduct social media campaigns to mold public opinion (Vu et al., 2021).

We must note here that we have omitted several entities and interactions, many of them less visible, and some of them illegitimate, that play important roles in the system. These include lobbying by unregistered organizations (such as those representing foreign governments, Qatargate being a recent example (Moller-Nielsen, 2023)), targeted political campaigns on social media that create filter bubbles and influence elections (such as the Cambridge Analytica scandal (Boldyreva et al., 2018)), and misinformation and disinformation campaigns such as those by climate-change deniers (Bloomfield & Tillery, 2019) and anti-vaxxers (Sufi et al., 2022). While some of these topics have been studied in prior work, others remain difficult to study due to a lack of availability of data. Even among the interactions we indicate in Figure 1.1, we do consider in this thesis those interactions that have been extensively studied in prior work, such as elections (Etter et al., 2014; Immer et al., 2020).

Although in an ideal scenario the system described would work to advance society's

interests, real-world instances have multiple issues that reduce their effectiveness.

Media outlets and Wikipedia editors have their own subjective biases that can creep into articles, even when there are well-formulated guidelines to ensure objectivity (Bennett, 2016; Greenstein & Zhu, 2012). This can cause citizens to use this biased information to make misguided decisions. An understanding of the different forms of biased language, and the use of tools that can identify, measure, and compare bias can mitigate this problem to some degree. For instance, citizens could use such tools to detect and ignore biased language and/or biased news sources, and editors could use them to reduce bias in articles prior to publication.

Public trust in institutions such as parliaments is the bedrock on which democracies function. Lobbying can play a constructive role in the democratic legislative process by helping law-makers be cognizant of the genuine needs of industries. However, lack of transparency in this process leads to a trust deficit between parliamentarians and the electorate and enables the self-interest of interest groups to potentially affect policy to the detriment of the wider societal interest (Solaiman, 2023).

Even though recent initiatives have attempted to enhance transparency by requiring lobbies to register publicly (European Union, 2011) and providing open data in the form of legislative amendments and parliamentary speeches (European Parliament, 2021a), it is difficult for an average citizen to study hundreds of speeches and amendments and find relationships between them and the policies mentioned in thousands of position papers published by multiple lobbies in their separate websites. Also, these policy documents often use domain-specific vocabulary and understanding them requires expert knowledge. The factors that influence a proposed amendment's inclusion into the final version of the law remain opaque to the general public. Analyzing these document datasets to bring out relationships between specific groups of lobbyists and parliamentarians and to derive insights into the factors that drive the legislative process can effectively enhance transparency and public trust, and thereby strengthen democracies.

Social media campaigns have the potential to increase public engagement on issues of societal interest such as climate change. However, their success has been limited due to the non-utilization of effective communication strategies based on research (Vu et al., 2021). Data regarding the 'likes' and 'retweets' of content on online social media such as Twitter[1] can be exploited to infer practical strategies to increase engagement.

We discussed several examples of problems that affect representative democracies and their potential solutions. These problems and solutions have traditionally been explored almost exclusively within the field of social science using manual methods and small sample sizes. Recently, however, the exponentially increasing amount of data available

[1]Twitter was recently renamed as X. However, as the name Twitter is known and used more widely in the literature at the time of writing, we use that name in the remainder of this thesis.

about such systems in digital form has necessitated the use of computational methods for its analysis, giving rise to the interdisciplinary field of computational social science (CSS). Going beyond traditional statistics, modern machine learning methods enable the extraction of insights from diverse datasets in multiple modalities.

As is clear from the foregoing description of social systems, a pervasive modality of social data is text in natural language, in the form of articles, social media posts, position papers, laws, etc. This is no accident: In fact, the ability to express abstract concepts through language is believed to be unique to humans and to be an essential skill that enabled us to collaborate and create socio-political systems in the first place (Harari, 2015). Analyzing text data is, therefore, imperative for understanding such systems comprehensively.

Methods for computationally analyzing text data have been studied within the field of natural language processing (NLP). Early NLP methods, starting in the 1950s, were symbolic and based on handwritten rules and theories of linguistics (Chomsky, 1965). These were dominant until the late 1980s, when statistical methods that learn models of language from text corpora, an approach called *corpus linguistics*, gained prominence. In the 2010s, neural networks were widely adopted for learning representations of words and sentences and for language modeling. At present, large language models based on deep neural networks with billions of parameters achieve state-of-the-art performance on nearly every NLP task, ranging from text classification to machine translation and question answering. And, in some cases, they even match or outperform humans.

However, though such deep models obtain a high level of performance, they are not easy to interpret. The input is processed through multiple steps of non-linear transformations, which makes it hard to determine the features that result in a particular output. Therefore, predictive models directly based on such deep architectures are not very useful for understanding the dynamics of the system being studied.

Hence, in this thesis, we seek to build *interpretable* text-based predictive models that can help us understand phenomena in the socio-political system of representative democracies. Guided by data availability and gaps in prior work, we decide to focus on four information flows, indicated by the solid arrows in Figure 1.1.

Specifically, we seek to answer the following research questions:

RQ1: Can we score, based only on their textual content, the subjective bias in web documents? Can we identify words that are indicative of bias?

RQ2: Can we assess the effectiveness of different communication strategies for social media campaigns by mining user engagement data?

RQ3: Can we discover links between parliamentarians and lobbyists by matching

their publically available text content, such as speeches and position papers?

RQ4: Can we predict the acceptance of edits to law proposals in parliamentary committees? What features of the edits and laws contribute to accepting edits?

We build our models by using learned representations of text semantics, topic models, and discrete choice models. We show that such models achieve a good level of predictive performance, without sacrificing interpretability. The model architecture is also simple enough that it can be generalized for studying other similar phenomena. In some cases, non-textual features, such as metadata, can give valuable additional information beyond that which is present in the text. Therefore, we include such features in our models for answering RQ2 and RQ4, which increases their accuracy and provides further insight.

In the following sections, we give an overview of the text representation methods, topic models, and discrete choice models we use in this work. This is not an exhaustive treatment of these topics; we give only the details necessary for understanding the methods used in subsequent chapters at the conceptual level. We expect the reader to have basic knowledge of machine learning, but no prior knowledge of NLP is assumed.

1.2 Distributional Representations of Text

Most machine learning methods cannot directly work with natural language text; first, it needs to be converted to a numerical representation (e.g., a vector of real numbers). This representation can then be used as features for machine learning models such as logistic regression or K-means clustering.

1.2.1 One-Hot Vectors

Assume that the words in the vocabulary are ordered (say by frequency in a text corpus). Each word is therefore associated with an index $k \in \mathbb{Z}_{\geq 0}$ corresponding to its position in the ordering. A simple method for representing texts numerically is to then represent each word by a sparse vector of the same size as the vocabulary, having a one at position k and zero elsewhere. A longer piece of text (such as a sentence, paragraph, or document) can then be represented as the sum or average of such vectors for each word in the text.

However, such a one-hot representation does not capture the meaning of a word. In particular, the representation vectors of synonyms are as far apart as those of completely unrelated words. This limits the generalization ability of downstream methods that use this representation. For instance, if a text classification model encounters a word that was not seen in the training set in a given piece of text, it cannot use the word for classification, even though its synonyms might have occurred in the training set.

1.2.2 Static Word Embeddings

This issue of limited generalization leads to the question of how can the meaning of a word be represented numerically. The *distributional hypothesis* in linguistics states that words that occur in similar contexts have a similar meaning (Harris, 1954). This is an idea that was summarized by the linguist J.R Firth as the famous statement "You shall know a word by the company it keeps!" (Firth, 1957). One natural way to capture this notion of meaning is to represent words by dense vectors such that the vectors of words that appear in similar contexts have a high vector similarity (e.g., cosine similarity).

The Word2Vec SkipGram algorithm introduced in Mikolov et al. (2013) was one of the first methods that used this hypothesis to efficiently learn word vector representations (word embeddings) in a self-supervised way from large text corpora. The algorithm uses two dense vectors to represent each word in the vocabulary: the *input vector* and the *context vector* (denoted as $\mathbf{v_w}$ and $\mathbf{v_w}'$, respectively). Given a large text corpus, the algorithm then essentially consists of moving a sliding window through the text and maximizes the probability of observing the words surrounding the word in the center of the window, as represented by their context vectors, given the word in the center, as represented by its input vector. For a given pair of center and context words (w_I, w_O), we thus maximize

$$P\left(w_O \mid w_I\right) = \frac{\exp\left(\mathbf{v'_{w_O}}^{\top} \mathbf{v_{w_I}}\right)}{\sum_{w \in \mathcal{V}} \exp\left(\mathbf{v'_w}^{\top} \mathbf{v_{w_I}}\right)}, \tag{1.1}$$

where $\mathcal{V}$ is the vocabulary.

The procedure is an instance of maximum likelihood estimation (MLE), where the likelihood of the parameters $\mathbf{v_w}$ and $\mathbf{v_w}'$ are maximized given the data in the form of pairs of the center word and context word as generated by the sliding window. Although non-convex, a local maximum of the likelihood can be found through stochastic gradient ascent. However, computing the (stochastic) gradient of the exact likelihood is computationally prohibitive, as it involves computing a sum over all words in $\mathcal{V}$ (the denominator in Equation 1.1).

To overcome this issue, Mikolov et al. (2013) propose a method called negative sampling (NS), a simplified version of noise contrastive estimation (NCE). NCE reframes the problem of maximizing the likelihood to a problem of binary classification, where the objective is to distinguish whether a sample (word pair (w_1, w_2)) comes from the true data distribution (w_2 appears in the context of w_1 in natural language text) or a known noise distribution (e.g., w_2 is uniformly sampled from $\mathcal{V}$). Mnih and Teh (2012) show that, as the ratio of noise samples to true observations increases, the NCE gradient

approaches the maximum likelihood gradient.

NS further simplifies NCE by defining the binary classification probabilities as logistic sigmoid functions of the parameters. Although this results in the loss of the asymptotic guarantee that the NCE gradient had, Mikolov et al. (2013) find that this does not degrade the quality of the learned vector representations. Unlike NCE, NS does not need the numerical probabilities of the noise distribution for computing the gradient. We then have the following loss function to minimize for a given (w_I, w_O) pair (instead of maximizing $P(w_O \mid w_I)$):

$$L(\mathbf{v}, \mathbf{v}') = -\left[\log \sigma\left({\mathbf{v}'_{\mathbf{w_O}}}^{\top} \mathbf{v}_{\mathbf{w_I}}\right) + \sum_{i=1}^{k} \mathbb{E}_{w_i \sim P_n(w)}\left[\log \sigma\left(-{\mathbf{v}'_{\mathbf{w_i}}}^{\top} \mathbf{v}_{\mathbf{w_I}}\right)\right]\right], \tag{1.2}$$

where k is the number of negative (noisy) samples[2] and $P_n(w)$ is the noise distribution. Mikolov et al. (2013) experiment with different choices of $P_n(w)$ and find that the unigram distribution raised to the 3/4th power performs best for several tasks such as solving analogies and language modeling.

Words paired with a similar set of true context words w_O (hence a similar set of context word vectors $\mathbf{v}'_{\mathbf{w_O}}$), have similar meanings according to the distributional hypothesis. Minimizing $L(\mathbf{v}, \mathbf{v}')$ tends to draw close the $\mathbf{v}_{\mathbf{w_I}}$ of such words and, as a result, they will have a high cosine similarity with each other. Once training is completed, the vectors $\mathbf{v}_{\mathbf{w_I}}$ constitute the word embeddings of all words in the vocabulary.

Given the word embeddings, a simple way to construct representations for longer pieces of text, such as sentences, is by averaging these embeddings for the words present in the text. This average vector can then be used as a feature vector for downstream tasks, such as text classification, by using models such as logistic regression. Although by simply averaging, we lose all information about the order of the words, it has been shown in the literature that it is a quite strong baseline for sentence representations, even outperforming more complex models based on neural networks (Arora et al., 2016; Mu et al., 2017).

The SkipGram algorithm of Mikolov et al. (2013) described above learns a distinct vector for each word. However, representing each word by a distinct vector without any parameter sharing makes it difficult to learn good representations of the rare words. For instance, the same word often has different inflected forms that reflect tense, number, gender, case, etc. (e.g., *kill*, *kills*, *killed*, *killing*), and some of these forms can be rare, especially in morphologically rich languages such as French or German. Furthermore, the meaning of some words is composed of the meaning of their sub-units, and words that

[2] k is a hyperparameter, usually set to 5 for large text corpora.

share sub-units can share some aspect of their semantics (e.g., *geothermal* and *geography*). The information in these sub-units that are shared between rare and more frequent words and forms could be exploited to learn better representations of the rare words and forms.

Bojanowski et al. (2017a) propose an improvement over the SkipGram algorithm that takes this approach by representing each character n-gram by a vector, with a word being represented by the sum of the vectors of the character n-grams present in it[3]. This sum replaces the vectors $\mathbf{v_{w_I}}$ in the SkipGram model; the rest of the model is identical. This formulation enables the representations of the character n-grams (which are the sub-units of words) to be shared across word forms, thus enabling better learning of the representations of rare words and word forms. We refer to this model as *fastText (Unsupervised)*, based on the name of the Python library (fastText) released by the authors for training and working with these vectors (Bojanowski et al., 2017b).

An additional important advantage of this method over SkipGram is that while using the pre-trained vectors for downstream applications, we can obtain reasonable representations of words that did not occur even once in the pre-training data (Out-Of-Vocabulary (OOV) words), if they share sub-units with words that did occur. Bojanowski et al. (2017a) demonstrate this quantitatively by using superior performance on word similarity tasks, especially when the size of the pre-training corpus is small (and OOV words are more frequent in the evaluation data). They also show this qualitatively by showing pairs of sub-units of OOV and non-OOV words whose embeddings have high similarities.

When a sufficiently large labeled dataset is available for training a classification model, learning the embeddings jointly with the model weights can help to capture in the embedding the aspects of word meaning that are relevant to the classification task. This can give better performance in some downstream tasks compared to using pre-trained embeddings learned in a self-supervised fashion on very large corpora (e.g., SkipGram, fastText (Unsupervised)).

For instance, the pre-trained self-supervised embeddings for the words *good* and *bad* have a fairly high cosine similarity[4], as most of the words that appear in their context are similar (most things described as good in one text in the pre-training corpus are also described as bad in another text in the corpus, e.g. *The food was good.*, *The food was bad.*). Even though these words are similar in the distributional sense, they are actually opposites of each other. Hence, using their pre-trained embeddings in a downstream task such as sentiment classification of product reviews can negatively affect the results. But if the embeddings were trained jointly with the sentiment classifier by using information from labels about the true sentiment of the review, the learned vectors for *good* and *bad* would be quite dissimilar because they are associated with different labels.

[3]The entire word is also considered as a character n-gram and associated with a vector of its own.

[4]For embeddings pre-trained on the English Wikipedia using fastText (Unsupervised) (Bojanowski et al., 2017c), the cosine similarity between *good* and *bad* is 0.67. For comparison, the similarity between *good* and *wonderful* in the same model is only 0.59.

Joulin et al. (2017) propose an algorithm for computing word embeddings with a supervised text classification objective. Their algorithm represents each word, character n-gram, and word n-gram[5] by a distinct embedding and computes the representation of a longer piece of text (such as a sentence) by the average of the embeddings of all its sub-units. It then jointly learns these embeddings and the weights of a linear classifier to minimize cross-entropy loss, given a labeled training set of texts. We refer to this algorithm as *fastText (Supervised)*, as its implementation is provided in the same fastText library mentioned earlier (Bojanowski et al., 2017b).

1.2.3 Dealing with Polysemy

All the aforementioned methods learn a single vector per word to represent its meaning. However, in many languages including English, the same word form can have multiple meanings, depending on the context in which it is used - a phenomenon called polysemy.

For instance, consider the word *bank*. In the sentence "John went to the bank to deposit money in his account.", *bank* refers to the financial institution, whereas in the sentence "John swam across the river to reach the opposite bank.", *bank* refers to the land bordering a river. In both sentences, the meaning is made clear by the context words - 'deposit', 'money', and 'account' in the first sentence, and 'swam', 'across', and 'river' in the second sentence.

If only a single vector is used to represent *bank* using SkipGram or fastText, it will encode a weighted average of the multiple meanings in different contexts, with meanings in more frequent contexts being attributed more weight. This could result in the loss of information about less frequent senses of the word. This could also decrease the embedding quality for the frequent senses if these are sufficiently far apart: *Bank* would need to be less similar to *financial* in order for it to be more similar to *river*, as *financial* and *river* do not have many context words in common.

Motivated by this issue, researchers have made several attempts to move beyond a single vector representation of words. One body of work extends the SkipGram model, enabling multiple vectors per word (one for each of its senses) and learning both the senses and vectors in a self-supervised manner, based on the difference in context word distributions (Bartunov et al., 2016; Neelakantan et al., 2014; Tian et al., 2014). Another line of work moves beyond point vectors and represents words as probability distributions, thus representing uncertainty or breadth in meaning by variance (distributions representing polysemous words have high variance) (Bražinskas et al., 2018; Vilnis & McCallum, 2014), or representing polysemous words by multimodal distributions with each mode corresponding to a different sense (Athiwaratkun & Wilson, 2017; Athiwaratkun et al., 2018).

[5]It is possible to omit the representations of character n-grams and/or word n-grams if so desired.

Although these methods outperform single vector representations on intrinsic evaluations in an artificial setting such as word similarity benchmarks, the observed performance improvement in real downstream applications has been mixed. J. Li and Jurafsky (2015) conduct an extensive evaluation comparing single-vector representations and multi-vector embeddings on a wide variety of tasks. They show that though the latter performs better on some of the tasks, this improvement disappears when using more complex neural network models for the tasks or when the dimensionality of the single vector representations is increased to have a fairer comparison in terms of parameters with the multi-vector models.

Here, we take a brief detour into language models, as it is a pre-requisite before introducing a more modern solution to the problem of polysemy, namely contextual word embeddings.

1.2.4 Language Models

A language model (LM) is a probabilistic model of sequences of words in a language. Given a sequence of words $S = (w_1, w_2, \ldots, w_T)$, a traditional LM computes the probability of S by using an auto-regressive factorization,

$$P(S) = P(w_1, w_2, \ldots, w_T) = P(w_1)P(w_2|w_1)\ldots P\left(w_T|\left(w_1, w_2, \ldots w_{T-1}\right)\right) \tag{1.3}$$

One of the simplest such LMs is an n-gram LM that makes a Markovian assumption to limit the past context used for modeling the future to a fixed length of $(n-1)$ words. For instance, a bigram LM makes the assumption that

$$P\left(w_t|\left(w_1, w_2, \ldots w_{t-1}\right)\right) = P(w_t|w_{t-1}), \quad \forall t. \tag{1.4}$$

The probabilities $P(w_t|w_{t-1})$ are estimated using the empirical word and bigram frequencies from a text corpus.

This approach is clearly not scalable to longer contexts. As n is increased, the number of possible n-grams and the probabilities to be estimated grow exponentially. Also, as longer n-grams occur less frequently and each n-gram is treated as distinct with no notion of semantic relatedness, we also need much larger text corpora to have reliable estimates of the probabilities.

One possible solution to this issue is to use a class of neural networks called recurrent neural networks (RNNs) to model word sequences. RNNs can model sequences of

arbitrary length and do not need to restrict to a fixed context size. They achieve this by parametrizing the probability $P\left(w_t|\left(w_1, w_2, \ldots w_{t-1}\right)\right)$ as

$$P\left(w_t|\left(w_1, w_2, \ldots w_{t-1}\right)\right) = P(w_t|\mathbf{h_t}), \tag{1.5}$$

where $\mathbf{h_t} \in \mathbb{R}^H$ is called the *hidden state* at time step t and H is the dimension of the hidden layer that is a hyperparameter of the model. $\mathbf{h_t}$ is computed as[6]

$$\mathbf{h_t} = f(\mathbf{h_{t-1}}, \mathbf{x_t}), \tag{1.6}$$

where $f(.)$ is a non-linear recurrence function that varies depending on the specific type of RNN and $\mathbf{x_t}$ is the input at time step t. The input is the embedding of word w_t in the sequence. $P(w_t|\mathbf{h_t})$ is typically parametrized as a linear transformation of $\mathbf{h_t}$, followed by the softmax function.

RNNs can be stacked vertically on top of each other to form multi-layer stacked RNNs (Schmidhuber, 1992). The inputs of the bottom RNN layer in the stacked RNN are the embeddings of the words in the sequence, while the inputs of subsequent layers are the hidden layer states $\mathbf{h'_t}$ of the RNN layer that is below this layer in the stack. Depending on the task for which the RNN is to be used and the amount of training data available, the word embeddings are either pre-trained using the methods seen earlier in Section 1.2.2 and kept constant (Qi et al., 2018), or trained jointly with the rest of the model's parameters.

All parameters are the same for every time-step of an RNN, which enables it to model sequences of arbitrary length. These networks are trained by finding, through gradient-based optimization methods, the values of the parameters that maximize the likelihood of sequences in a training set. The network learns to encode relevant information about all of the past context in the hidden state $\mathbf{h_t}$.

In theory, RNNs should be able to model arbitrarily long contexts. But, in practice, 'vanilla' RNNs, which parameterize $f(.)$ as a single linear transformation followed by a non-linearity, struggle to learn long-range dependencies. This occurs due to the *vanishing gradient problem* (Bengio et al., 1994), whereby the norm of gradient-based updates for the parameters become vanishingly small, as they are backpropagated through multiple steps of recurrence. More complex parametrizations of $f(.)$, such as long short-term memory (LSTM) (Hochreiter & Schmidhuber, 1997) and gated recurrent unit (GRU)

[6]The initial hidden state $\mathbf{h_1}$ is set arbitrarily, for instance to an all-zeros vector.

(Chung et al., 2014), mitigate this issue by using neural 'gates' to control information flow outside non-linearities.

As LMs are probabilistic models of word sequences, we can sample from them to complete partial sequences or generate new word sequences. Neural LMs can also be used to build sequence-to-sequence models (Sutskever et al., 2014), where one LM acts as the 'encoder' that takes a sequence as input and encodes its information into hidden states that are then used by a 'decoder' LM to generate the output sequence. Such models find applications in tasks such as paraphrasing, summarization, and machine translation.

In the sequence-to-sequence model proposed by Sutskever et al. (2014), the encoder LM represents all information about the input sequence in its final hidden state; this information is then set to be the decoder LM's first hidden state. This introduces a bottleneck in the information flow from the encoder to the decoder, as the encoder's final hidden state has a fixed size independent of the length of the input sequence. In fact, Cho et al. (2014) find that the performance of encoder-decoder models rapidly degrades as the lengths of input sequences increase.

Furthermore, when the input sequence is long, the first hidden states of the decoder, which are responsible for generating the first tokens of the output sequence, are several time steps away from the first hidden states of the encoder that contain the most information about the first tokens of the input sequence. This is a disadvantage in many sequence-to-sequence tasks, including machine translation, where the first tokens of the output sequence are closely related to the first tokens of the input sequence[7]. Indeed, Sutskever et al. (2014) find that the performance of their model improves when the input sequence is reversed relative to the output sequence.

Bahdanau et al. (2015) propose an elegant solution to both problems by introducing a mechanism called *attention*: it endows the decoder with the ability to look up or 'attend' to, in each time step, the most relevant hidden states of the encoder for the token it needs to generate at that time step. This enables the encoder to use all of its hidden states for representing the input sequence (instead of just the final state as was the case in Sutskever et al. (2014)), thereby eliminating the bottleneck. This also enables the decoder to effectively access information from any hidden state of the encoder, irrespective of the number of recurrence steps between them, thus enabling the model to learn long-range dependencies. Here, we give some more details about attention as it is a key innovation that gave rise to many new models (some of which we will discuss later) that achieved new state-of-the-art results in nearly every task in NLP, and even in tasks in other domains of artificial intelligence (AI) such as computer vision (CV).

In vanilla encoder-decoder models, such as the one in Sutskever et al. (2014), the output sequence is modeled as follows:

[7] Except in cases where the input and output language have opposite word order.

$$P(S') = \prod_{t'=1}^{T'} P\left(w'_{t'} | \left(w'_1, \ldots, w'_{t'-1}\right), \mathbf{c}\right), \tag{1.7}$$

where $S' = (w'_1, \ldots, w'_{T'})$ is the output sequence and $\mathbf{c}$ is the fixed-size context vector that encodes the information about the input sequence ($\mathbf{c}$ is the final hidden state of the encoder in the model of Sutskever et al. (2014)). Note that this conditioning on $\mathbf{c}$ distinguishes the decoder from a standard LM (Equation 1.3). Bahdanau et al. (2015) proposes to modify this to

$$P(S') = \prod_{t'=1}^{T'} P\left(w'_{t'} | \left(w'_1, \ldots, w'_{t'-1}\right), \mathbf{c_{t'}}\right), \tag{1.8}$$

where the context vector $\mathbf{c_{t'}}$ is different for each time step t'. The context vector is calculated as a weighted sum of the encoder hidden states $\mathbf{h_t}$, i.e. $\mathbf{c_{t'}} = \sum_{t=1}^{T} \alpha_{t't} \mathbf{h_t}$. The weights $\alpha_{t't}$ are called the *attention weights*, and they are computed as

$$\alpha_{t't} = \frac{\exp\left(s_{t't}\right)}{\sum_{t=1}^{T} \exp\left(s_{t't}\right)}, \tag{1.9}$$

where $s_{t't} = r(\mathbf{q_{t'}}, \mathbf{k_t})$ is the *relevance score* between the *query vector* $\mathbf{q_{t'}}$ at decoder step t' and the *key vector* $\mathbf{k_t}$ at encoder step t. Bahdanau et al. (2015) define $r(.)$ to be a feed-forward neural network, $\mathbf{q_{t'}}$ as the decoder hidden state at step $(t'-1)$, and $\mathbf{k_t}$ as $\mathbf{h_t}$[8]. As a result, attention can be viewed as a look-up mechanism where the decoder uses a query ($\mathbf{q_{t'}}$) to obtain relevant values ($\mathbf{h_t}$) based on their associated keys ($\mathbf{k_t}$).

In addition to improving performance, attention also provides explainability. For the task of machine translation studied in Bahdanau et al. (2015), by using attention weights computed by the trained model, we can visualize the alignment of words with similar meanings in the input and output language.

However, RNN LMs, including those with attention, have two major issues. The computation in RNN LMs is inherently sequential (see Equation 1.6, $\mathbf{h_t}$ can be computed only after $\mathbf{h_{t-1}}$). This prevents processing parts of the word sequence in parallel when the entire sequence is known (for instance, this is the case for the encoder in an encoder-decoder translation model), thereby reducing efficiency. Additionally, even with

[8] In general, in attention mechanisms, $\mathbf{k_t}$ is defined as a linear transformation of $\mathbf{h_t}$. For instance, see Vaswani et al. (2017).

modifications such as LSTMs, RNNs still struggle to learn long-range dependencies, as the information between any two words in the sequence needs to flow through all hidden states that correspond to the intermediate words.

In their landmark paper, Vaswani et al. (2017) propose a new architecture called the Transformer that solves these issues. Transformer replaces recurrence with *self-attention*, whereby words in a sequence attend to other words within the same sequence in order to construct its hidden state. This enables the hidden states of all words in a sequence to be computed in parallel, hence improving efficiency. Furthermore, self-attention enables information to flow directly between any two words in the sequence, without passing through representations of intermediate words, thus enabling the model to effectively learn long-range dependencies. In order to enable the model to make use of information about the word order, a *positional encoding* (either fixed or learned) is added to the input representations of the words[9]. Vaswani et al. (2017) show that their Transformer-based encoder-decoder model significantly outperforms RNN-based models on machine translation and incurs only a fraction of the training cost.

Beyond machine translation, the Transformer model and its variants are used within NLP, CV, and other domains of AI, and they achieve state-of-the-art results on a wide variety of tasks. In fact, GPT-4 (OpenAI, 2023b), LLaMA (Touvron et al., 2023), and related Large LMs (LLMs) that are behind chatbots such as ChatGPT (OpenAI, 2022) and others that have recently received much public attention, are based on this architecture.

Contextual Word Embeddings

After the detour into language models, we now get back to word embeddings. In Section 1.2.3, we described word embedding models that attempted to account for polysemy by learning a *finite set* of vectors per word, one for each of its senses. We now describe a different approach that uses pre-trained neural LMs to construct representations of words at the *token* level, i.e., each occurrence of the word has a different vector that is computed as a function of its context (a *contextual embedding*, as opposed to the *static embeddings* seen before (e.g., SkipGram, fastText)). Such an approach enables us to capture even fine-grained differences in the meaning of the word in the possibly *infinite* number of contextual embeddings.

Peters et al. (2018) propose one of the first widely used models that follow this approach, called embeddings from language models (ELMo). ELMo consists of two LMs that model a sequence in the forward and backward directions (together they form a biLM). The hidden state of the biLM for each time step of the sequence is defined as the concatenation of the states of the forward and backward LM at that time step. ELMo uses vertically stacked (multi-layered) LSTM RNNs for each of the LMs. The hidden state of the biLM

[9] RNNs do not require this as the computation is sequential and follows the order of the word sequence.

corresponding to each word is defined as its contextual embedding[10]. Remarkably, Peters et al. (2018) show that modifying existing baseline models for various NLP tasks by simply adding ELMo embeddings enables them to perform better than the state of the art.

Instead of using RNNs, Devlin et al. (2019) use the Transformer architecture that we previously described to obtain contextual embeddings; it is called Bidirectional Encoder Representations from Transformers (BERT)[11]. BERT is pre-trained with two objectives: masked language modeling (MLM) and next-sentence prediction (NSP).

Whereas traditional LMs (Equation 1.3) are autoregressive, hence unidirectional, MLMs are bidirectional. Inspired by the *Cloze* procedure for measuring the readability of texts (Taylor, 1953), MLMs are trained by randomly replacing some of the tokens[12] in a sequence with a special '[MASK]' token and predicting these tokens from the entire sequence. Although they are not autoregressive, MLMs can still model the probability of a sequence (like traditional LMs) when interpreted as Markov random fields (A. Wang & Cho, 2019).

NSP is a simple binary classification task of predicting whether a given pair of sentences appear next to each other in a corpus or not. A special '[CLS]' token is prepended to the concatenation of the two sequences, and the hidden state corresponding to this is used for the classification.

After pre-training, the contextual embedding for a word is obtained from the hidden states corresponding to its token[13]. Devlin et al. (2019) show that adding a simple linear classifier on top of the BERT hidden states and jointly training the classifier weights and fine-tuning the BERT parameters achieve a new state-of-the-art performance in eleven different NLP tasks. Adding the pre-trained BERT embeddings, without fine-tuning directly to a task-specific architecture (a similar approach to ELMo), also achieves nearly the same performance.

In the original Transformer encoder architecture (on which BERT is based), every token attends to every other token in the sequence. Hence, the computational complexity of this self-attention operation grows quadratically with the sequence length. This makes it too expensive for modeling sequences that are more than a few hundred tokens in length, such as Wikipedia articles or other documents[14]. Although the document could be split into chunks and each chunk modeled separately, this loses information across chunks.

[10] Strictly speaking, ELMo uses a linear combination of the hidden states from different layers of the biLM.

[11] BERT uses the Transformer encoder, where words can attend to other words that are both before and after it, hence bidirectional.

[12] The tokens here are subword units, specifically wordpieces (Wu et al., 2016).

[13] An average is taken if the word is split into multiple subword tokens.

[14] The maximum sequence length of the pre-trained BERT model in Devlin et al. (2019) is 512 tokens.

Beltagy et al. (2020) propose a modified Transformer architecture called the Longformer; it has an attention mechanism that scales linearly with the sequence length, thus enabling it to process much longer sequences[15]. The attention mechanism is a combination of local and global attention. Local attention is achieved through a sliding window, where the hidden state of each token in a layer attends to a fixed-size window of hidden states of the tokens to the left and right of it, in the layer below. When there are multiple vertically stacked hidden states, as is usually the case for the Transformer architecture, the states in the top layer are able to have a large receptive field. In addition to this local attention, for some tasks, the Longformer also incorporates a global attention mechanism where a few task-specific tokens attend to all tokens in the sequence (and vice-versa). Beltagy et al. (2020) show that this architecture consistently outperforms the base Transformer (that breaks a long sequence into chunks and processes each individually) for long-document tasks including document classification and question answering.

1.2.5 Sentence Embeddings

As we mentioned previously in Section 1.2.2, averaging the static embeddings for the words present in a sentence is a strong baseline method for embedding the sentence. However, this does not necessarily hold true for averaging contextual embeddings. In fact, Reimers and Gurevych (2019a) show that averaging the contextual embeddings produced by BERT results in a performance worse than averaging static embeddings for semantic textual similarity (STS) tasks. STS is analogous to the word similarity benchmarks used for word embeddings; the cosine similarity of the embeddings of two sentences is compared to human similarity judgments.

Reimers and Gurevych (2019a) propose a method, called Sentence BERT (SBERT), for making the BERT contextual embeddings better suited for obtaining sentence embeddings. Their method consists of fine-tuning the pre-trained BERT model for a three-way sentence-pair classification task called natural language inference (NLI), where given a pair of sentences (A, B), the model has to determine the inference relation between them: *A entails B*, *A contradicts B*, or neither (*neutral*). Two widely used datasets that contain such sentence pairs are the Stanford NLI (SNLI) dataset (Bowman et al., 2015), which contains 570K pairs, and the Multi-Genre NLI (MNLI) dataset (Williams et al., 2018), which contains 433K pairs from more diverse sources. Reimers and Gurevych (2019a) use a combination of these datasets for fine-tuning. Example sentence pairs from SNLI with their inference labels are given in Table 1.1.

For fine-tuning BERT for NLI, they use the 'Siamese' network architecture (see Figure 1.2) where the sentences in the pair are fed to two BERT networks whose weights are shared. The averages of the contextual embeddings from each BERT (denoted $\mathbf{u}$ and $\mathbf{v}$ in Figure 1.2) are concatenated, along with their element-wise absolute difference $|\mathbf{u} - \mathbf{v}|$

[15]The pre-trained models released by Beltagy et al. (2020) can process upto 4,096 tokens.

Table 1.1: Example sentence pairs and labels from the SNLI dataset (Bowman et al., 2015).

Sentence A	**Sentence B**	**Label**
A soccer game with multiple males playing.	Some men are playing a sport.	Entailment
A black race car starts up in front of a crowd of people.	A man is driving down a lonely road.	Contradiction
An older and younger man smiling.	Two men are smiling and laughing at the cats playing on the floor.	Neutral

that is then fed to a softmax classifier for the three-way classification.

Once the BERT weights are fine-tuned, the average of contextual embeddings constructed for a given sentence defines its SBERT embedding. Semantically similar sentences have a high cosine similarity between their SBERT embeddings. Reimers and Gurevych (2019a) show that SBERT outperforms averaged static word embeddings and other approaches for sentence embedding using neural networks (other than BERT) such as InferSent (Conneau et al., 2017) and Universal Sentence Encoder (Cer et al., 2018), in the STS task.

The authors also fine-tune different models on other sentence-pair tasks besides NLI and on triplet comparison tasks by making small and straightforward modifications to the architecture. We do not describe these for the sake of clarity in presentation. They release all the fine-tuned models through their Python package for working with SBERT called Sentence-Transformers (Reimers & Gurevych, 2019b).

Fine-tuned SBERT embeddings are a good choice for 'off-the-shelf' sentence embeddings when scoring semantic similarity in the unsupervised setting for tasks such as search or clustering. However, for other tasks such as classification, when sufficient labeled data is available, a better approach is usually to fine-tune BERT from scratch, as the embeddings are tailored for that specific task.

It is also worth noting that SBERT is *not* the best method if the performance on the sentence-pair task (e.g., NLI) is of importance. Directly feeding the sentence-pair (A, B) to a single BERT model (the so-called *cross-encoder* approach), which is then fine-tuned for that task, usually performs better. This is probably because SBERT might not be able to include all relevant information about sentence A in its embedding, as the information that is relevant could depend on sentence B (e.g., some phrase in A could be negated in B). The cross-encoder has an advantage here as it can use inter-sentence ('cross') attention, which is absent in the case of SBERT, for performing the task.

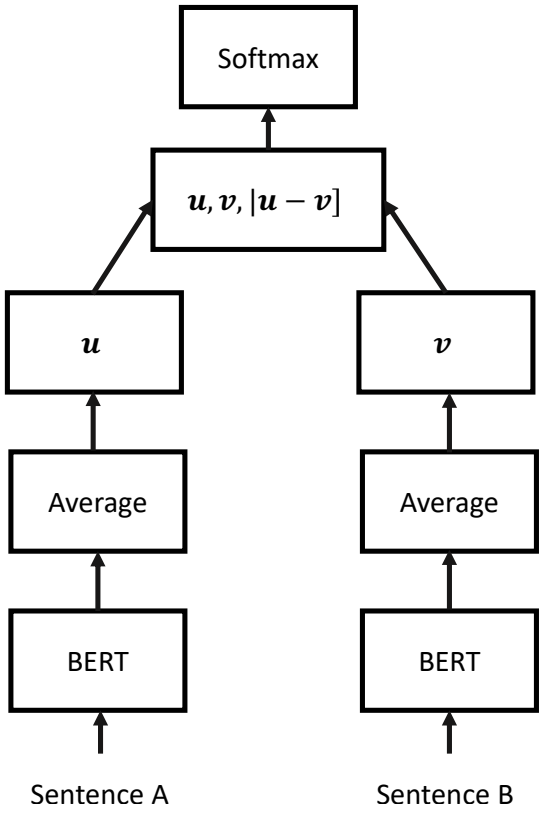

Figure 1.2: Architecture of SBERT at training time. Figure adapted from Reimers and Gurevych (2019a).

However, in the cross-encoder approach, the model does not learn to generate sentence embeddings. Consequently, model inference needs to be done for every sentence pair that needs to be compared. This is computationally expensive, especially for tasks that (such as clustering and semantic search) require numerous comparisons.

The advantage of SBERT is that it learns to produce embeddings that can then be compared by a simple cosine similarity computation. Thus, it is computationally much less expensive than the cross-encoder for moderately large datasets (say, a few thousand sentences), because the embeddings for every sentence can be pre-computed and stored and the subsequent comparisons are much faster than running the cross-encoder for every pair.

Finally, it is worth noting that, although the SBERT models were trained to embed sentences, they work nearly as well for representing slightly longer pieces of text such as a few sentences or a short paragraph.

1.3 Topic Models

In the previous section, we described methods inspired by the distributional hypothesis for representing words and longer individual pieces of text by dense vectors capturing their semantics (*embeddings*). However, although such methods enable state-of-the-art performance on many downstream NLP tasks, the representations themselves are not interpretable, other than by their relation to other representations. For instance, the embeddings of two semantically similar words or sentences have a high cosine similarity, but it is not clear what each dimension of an embedding vector represents.

When the corpus is organized into *documents* that are semantically coherent, there is an alternative approach to model it, namely by means of latent *topics*. Topics are defined in relation to words[16] and the vector of topic weights corresponding to a document provides an interpretable representation of its semantics. Topics and document-topic weights are jointly learned from the corpus in an unsupervised manner, whereas the number of topics is typically pre-defined as a hyperparameter.

One of the first methods to follow this approach is latent semantic analysis (LSA) (Deerwester et al., 1990). In LSA, the corpus is first represented as a term-document matrix $\mathbf{X} \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{D}|}$, where $|\mathcal{V}|$ is the number of terms (words or n-grams) in the vocabulary and $|\mathcal{D}|$ is the number of documents in the corpus. Each element X_{wd} is the term frequency-inverse document frequency (TF-IDF) score of the term-document pair (w, d), defined as[17]

$$X_{wd} = \mathrm{tf}(w, d) \times \mathrm{idf}(w) = \mathrm{tf}(w, d) \times \log\left(\frac{|\mathcal{D}|}{|d \in \mathcal{D} : \mathrm{tf}(w, d) > 0|}\right), \tag{1.10}$$

where $\mathrm{tf}(w, d)$ is the frequency of occurrence of term w in d, normalised by the frequency of the most frequent term in d. Intuitively, X_{wd} captures the importance of the term w for determining the topic of document d, which depends on both $\mathrm{tf}(w, d)$ and $\mathrm{idf}(w)$; the latter being a measure of the specificity of the term w (words that occur in nearly every document such as *the*, *of* etc. will have low $\mathrm{idf}(w)$). LSA then consists of performing the truncated singular value decomposition (SVD) of $\mathbf{X}$, keeping only the top K singular values, where K is a hyperparameter that corresponds to the number of topics. We have,

$$\mathbf{X} \approx \widetilde{\mathbf{X}} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathbf{T}}, \tag{1.11}$$

[16] More precisely, they are defined as vectors of the same size as the vocabulary where the vector component is high for the words associated with that topic.

[17] This is one of the many different variants of TF-IDF.

where $\mathbf{U} \in \mathbb{R}^{|\mathcal{V}| \times K}$ is the term-topic matrix, $\mathbf{V} \in \mathbb{R}^{|\mathcal{D}| \times K}$ is the document-topic matrix, and $\mathbf{S} \in \mathbb{R}^{K \times K}$ is the diagonal matrix containing the top K singular values along its diagonal. The columns of $\mathbf{U}$ are the topic vectors, the rows of $\mathbf{V}$ are the document-topic weights, and the singular values represent the importance of the topics in terms of variance captured.

However, LSA is not based on any underlying probabilistic model of the documents, which makes it difficult to interpret. For instance, it is unclear how to label a discovered topic if the relevant topic vector includes negative components. As a solution to this issue, Hofmann (1999) propose a probabilistic version of LSA, called probabilistic LSA (pLSA).

pLSA posits the following generation procedure for a given text corpus. First, choose a document $d \in \{1, 2, \ldots, |\mathcal{D}|\}$ with probability $P(d)$. Second, choose a topic $z \in \{1, 2, \ldots, K\}$ with probability $P(z|d)$. Finally, generate a new word $w \in \mathcal{V}$ for d with probability $P(w|z)$. The process is repeated until the corpus is fully generated. The model assumes that every pair (w, d) is generated independently and that the generation of the word w is conditionally independent of the document identity d, given the topic z.

This gives the joint probability model for a pair (w, d),

$$P(w, d) = P(d) \sum_z P(z|d)P(w|z) = \sum_z P(z)P(w|z)P(d|z). \tag{1.12}$$

The probabilities $P(z)$, $P(w|z)$ and $P(d|z)$ are the parameters of the model, and they are learned by maximum likelihood estimation by using the expectation maximization (EM) algorithm. The topic vectors are then given by $P(w|z)$ and the document-topic weights are given by $P(z|d)$, both of which are easily interpretable based on their role in the generation procedure.

Although this is an improvement over LSA, the generative model of pLSA is incomplete in the sense that there is no generation procedure for the probabilities $P(z|d)$. The learned model cannot be used to generate a new document that was not part of the text corpus used for learning the parameters (the *training corpus*). There is no natural way[18] to estimate the topic weights $P(z|d)$ for a given document that was not part of the training corpus.

Blei et al. (2003) propose a Bayesian extension of pLSA, called latent Dirichlet allocation

[18]Hofmann (1999) propose a heuristic called 'folding-in' for this purpose, but this method does not follow naturally from the model definition and makes the determination of topic-weights inconsistent for documents inside and outside the training corpus. Also, Blei et al. (2003) empirically show the superiority of LDA's approach in downstream tasks.

(LDA); it solves this issue. LDA posits that the topic weights for any document d, $P(z|d)$, are sampled from a K-dimensional Dirichlet distribution with parameter vector α. The model parameters (α and the word probabilities $P(w|z)$) are learned by using variational inference (Jordan et al., 1999) to find a tight lower bound to the log-likelihood of the data; this bound is then maximized with respect to the parameters.

Once the model is learned, a new document can be generated by sampling the topic weights $P(z|d)$ for the document from the Dirichlet distribution and then proceeding as in the case of pLSA to generate the words of the document. The document-topic weights $P(z|d)$ for a given document d, including documents not part of the training corpus, can be estimated by using the posterior distribution[19] for $P(z|d)$ given the words in the document and the learned parameters α and $P(w|z)$.

The models discussed so far, including LDA, rely only on the statistics of word counts in documents used for topic modeling, treating each word as distinct with no prior notion of semantic relatedness between them. This makes it difficult to process corpora with large vocabularies as shown by Dieng et al. (2020). To solve this issue, they propose the embedded topic model (ETM) that combines LDA with word embeddings.

In ETM, each topic and word has a dense vector embedding. And the word probabilities of the topics, $P(w|z)$, are parametrized in terms of the dot-product between the topic and word embedding. The word embeddings can be pre-trained (using SkipGram, for instance). This enables the model to make use of word similarity information present in the embeddings, to generalize better, and to infer topics, even for the words that were absent in the training documents for topic modeling. Dieng et al. (2020) demonstrate the superior performance of ETM, compared to LDA and similar competing approaches, in terms of topic interpretability and predictive power, especially in cases where the vocabulary size is large.

The dominant paradigm for topic discovery has been to use probabilistic generative models that aim to maximize the likelihood of the training corpus (e.g., pLSI, LDA, ETM). However, the primary use of topic models, especially in the context of this thesis, is to extract interpretable topics that can be used as document features for downstream tasks. Chang et al. (2009) show that models that maximize the likelihood do not always give the most interpretable topics. Recently, alternative approaches based on clustering have been proposed for topic discovery. They are conceptually much simpler and computationally less expensive, and they match or outperform the generative models in terms of topic interpretability.

Sia et al. (2020) show that centroid-based clustering of pre-trained word embeddings can give topics of quality comparable to LDA (in terms of interpretability), with much lower computational cost. Each cluster centroid corresponds to a different topic, and the

[19] In practice, one takes the expected value of the variational approximation to the posterior.

relevance of a word to a topic is given by the proximity of the word embedding to the cluster centroid.

Grootendorst (2022) propose a model called BERTopic; it clusters documents based on their SBERT representations (the document is truncated to the maximum sequence length allowed by SBERT). Each cluster corresponds to a topic, and the relevance of a word to a topic is given by a modified version of its TF-IDF score that is computed after concatenating the documents in each cluster. Note that clustering SBERT representations of documents enables them to go beyond the 'bag-of-words' assumption that is commonly made in topic models and where the order of words within a document is not important for determining topics. Z. Zhang et al. (2022) perform a comprehensive evaluation of these and other clustering-based algorithms, and they show that they match or outperform LDA and more complex variants of it that use embeddings and neural networks, in terms of quantitative measures of interpretability such as topic coherence (similarity of words within a topic)[20] and topic diversity (difference between topics).

Meng et al. (2022) study the properties of the contextual word embeddings generated by BERT and by related models. They show that the embeddings form a large number of very small clusters (few words in size) based on fine-grained similarity instead of a small number of large clusters based on topical similarity. This suggests that a direct clustering in this space, by using a K-means style algorithm with a small K (corresponding to the number of topics), is likely to give poor-quality clusters. The space is also relatively high dimensional (768 dimensional in the case of BERT). This leads to issues for clustering, due to the so-called 'curse of dimensionality' (Beyer et al., 1999), whereby distance functions become less meaningful in high dimensional spaces (the distance to the farthest data point approaches the distance to the closest data point).

BERTopic mitigates this issue, to some extent, by first projecting the SBERT representations to a lower dimensional space by using UMAP (McInnes et al., 2018) before clustering. However, this does not directly address the problem of fine-grained clusters, as the projection only tries to maintain the structure of points lying on a manifold in the original space and does not take the clustering objective into consideration.

Meng et al. (2022) propose a model called TopClus that overcomes this issue by learning a dense embedding for each of the K topics (similar to ETM) on a lower dimensional latent space better suited for clustering than the BERT embedding space. It constructs document embeddings by an attention-based weighted average of BERT word embeddings and projects the word and document embeddings onto the latent space. It is important to note that the functions used to project between the BERT embedding space and the latent space are learned so that the projected word embeddings form K well-separated soft clusters in the latent space.

[20] Mimno et al. (2011) show topic coherence to be a reliable automatic quantitative measure of topic interpretability.

Table 1.2: Table of top words identified for the 'sports' topic on a news dataset by different topic modeling algorithms. Words not belonging to the topic are italicized. Table adapted from Meng et al. (2022).

LDA	**ETM**	**BERTopic**	**TopClus**
olympic	olympic	swimming	athletes
year	league	freestyle	medalist
said	*national*	*popov*	olympics
games	basketball	gold	tournament
team	athletes	olympic	quarterfinal

In addition to the clustering objective, the training procedure also ensures that the document embeddings can be reconstructed from the topic embeddings and that the word embeddings can be reconstructed from the projected word embeddings. Once the model is trained, it is straightforward to obtain the word probabilities for topics (using the proximity between word and topic embeddings in the latent space) and the document-topic weights (using the proximity between document and topic embeddings in the latent space). Meng et al. (2022) show that TopClus significantly and consistently outperforms LDA, ETM, and BERTopic on topic interpretability metrics including topic coherence and topic diversity. For a qualitative illustration, Table 1.2 lists the top words identified by these different algorithms for the 'sports' topic on a news dataset[21]; it is clear that TopClus achieves the most coherent topic description. More examples of topics and datasets are provided in Meng et al. (2022).

1.4 Discrete Choice Models

Modeling social phenomena often requires the estimation of human preferences and subjective quantities. Considering the phenomena we study in this thesis, the acceptance of edits and engagement with tweets are directly related to human preferences, whereas bias in web documents is a subjective quantity. Discrete choice models are the statistical tools of choice[22] in this setting. As a result, they inspire the models we build in subsequent chapters.

One of the first descriptions of a model of this nature was proposed by Zermelo (1929) in the context of estimating the relative skill of chess players from the outcomes of matches between pairs. The model is still used today as part of the Elo rating system (Elo, 1978).

In this model, each player i is associated with a skill parameter $p_i \in \mathbb{R}_{>0}$. The probability that i wins over j in a match is defined as

[21] For each algorithm, among the topics identified by it, the one that best represents 'sports' is chosen.
[22] No pun intended.

$$P(i \succ j) = \frac{p_i}{p_i + p_j}. \tag{1.13}$$

The parameters p_i are learned by maximizing their likelihood, given a dataset of match outcomes. Note that the parameters can be estimated only up to a multiplicative factor.

This pairwise comparison model was also proposed independently by Bradley and Terry (1952). They propose a reparametrization of the model as

$$P(i \succ j) = \frac{\exp(s_i)}{\exp(s_i) + \exp(s_j)}, \tag{1.14}$$

where the parameters $s_i \in \mathbb{R}$ can be optimised without constraint.

We can further parametrize s_i in terms of the *features* of the items being compared (say the chess players). Suppose each player is associated with a d-dimensional feature vector $\mathbf{x_i} \in \mathbb{R}^d$. For instance, this could be a one-hot vector representing the player's nationality. We could then let $s_i = \mathbf{w}^T \mathbf{x_i}$, where $\mathbf{w}$ is a learned parameter vector. The components of $\mathbf{w}$ after training could be interpreted to understand the effect of the features (in this example, the nationality) on the skill.

This parametrization allows for improved efficiency and generalization when the number of items N (here, the number of players) is large, but the items' quality that is being compared (here, the skill) can be represented in terms of a low-dimensional feature vector (here the nationality vector) with dimension $d \ll N$. Text features can be expressed as low-dimensional vectors by using embeddings or topic weights, as discussed in previous sections (in the current example, text features could include, for instance, Wikipedia articles on the players).

The Bradley-Terry model deals only with pairwise comparisons. To efficiently extend it to multi-way comparisons, we need to make use of the *choice axiom* proposed by Luce (1959) in the context of an individual choosing an item from a set of items. This axiom, also known as the independence from irrelevant alternatives (IIA), states that the ratio of the probabilities of choosing items i and j from a set of items $\mathcal{I}$ is independent of other items in the set, i.e.,

$$\frac{P(i \succ \mathcal{I} - \{i\})}{P(j \succ \mathcal{I} - \{j\})} = \frac{P(i \succ j)}{P(j \succ i)}, \quad \forall i, j \in \mathcal{I} \tag{1.15}$$

where $P(i \succ \mathcal{I} - \{i\})$ denotes the probability of choosing item i over the other items in $\mathcal{I}$. This is a restrictive assumption that does not hold in general. But, Luce (1959) argues that it could be reasonably expected to hold when the set $\mathcal{I}$ is not too large and cannot be decomposed into subsets of similar items.

Importantly, as shown by Luce (1959), the IIA assumption enables us to extend the Bradley-Terry model to the multinomial logit model that can deal with multiway comparisons. When parametrized in terms of features, this model is defined as

$$P(i \succ \mathcal{I} - \{i\}) = \frac{\exp\left(\mathbf{w}^T \mathbf{x_i}\right)}{\sum_{j \in \mathcal{I}} \exp\left(\mathbf{w}^T \mathbf{x_j}\right)}. \tag{1.16}$$

1.5 Outline and Contributions

In this thesis, we seek to understand social phenomena in representative democracies, through customized computational methods using data in the form of text in natural language. Our contributions towards answering the research questions in Section 1.1 fall into three main categories:

- **Datasets:** We curate several new datasets, including biased articles from *Wikipedia*, *Tweets* related to climate change, *position papers* of lobbies and *speeches* of parliamentarians.
- **Methods:** We develop *simple*, *efficient*, and *interpretable* methods for tasks, such as bias scoring and engagement prediction, by using text data, that could be applied at scale and generalized to related tasks.
- **Interpretations:** We demonstrate how our models can be interpreted to derive useful *insights* about the social phenomena being studied.

Here, we provide an outline of the remainder of the thesis.

First, in Chapter 2, we study subjective bias in the textual information on the Web, such as Wikipedia and news articles. We discuss the need for modeling bias in a *relative* sense instead of an absolute classification into biased and unbiased text. We curate a dataset of pairs of consecutive versions of the same Wikipedia article, where one version in the pair is more biased than the other. We use this dataset to train a Bradley-Terry model for comparing the versions in terms of bias, where the *bias score* of a version is defined as a function of its text features. We show that we can interpret the parameters of the trained model to discover the words most indicative of bias. We also demonstrate the generalizability of the model by applying it in many settings different from the training

domain - comparing bias in different Wikipedia articles, studying the temporal evolution of bias over the entire history of a Wikipedia article, comparing news sources based on bias, and scoring bias in law amendments, speeches, and tweets.

In Chapter 3, we study the influence of text content on user engagement in social media campaigns, considering the case of communication on Twitter about climate change as an example. We discuss the challenges of undertaking such a study, in particular, due to the presence of confounding factors that influence engagement such as an author's popularity. We are able to reduce the influence of such confounding factors by framing the problem as one of comparing the engagement of a pair of tweets from the same author. We then train a Bradley-Terry model for this task, using the topic weights of the tweets (extracted using TopClus) as text features, and using metadata information about the tweet as additional features. We interpret the trained model to discover the topics that contribute to high engagement and provide recommendations for optimizing communication about climate change in light of the findings.

Then, in Chapter 4, we shift our focus to a different part of the democratic system, specifically the influence of interest groups (lobbies) on lawmakers. We curate an extensive dataset of 48,970 position papers published by 2,558 registered lobbies by crawling each of their websites and classifying the PDF documents obtained. We also curate a dataset of 51,432 speeches made by the 849 members of the European Parliament (MEPs) present during the eighth term of the legislature (2014-2019). By comparing the position papers and speeches, on the basis of semantic similarity and entailment, we are able to discover interpretable links between lobbies and MEPs. In the absence of a ground-truth dataset of such links, we perform an indirect validation of the discovered links against a dataset, which we curate, of retweet links between the MEPs and lobbies and with the publicly disclosed meetings of MEPs. We also perform an aggregate analysis of the discovered links and show that the findings correspond to the expectations from the ideology of the MEPs' political groups (e.g., center-left groups are associated with social causes).

Finally, in Chapter 5, we look at the law-making process within the European Parliament (EP). We briefly describe the ordinary legislative procedure, whereby the European Commission (the executive) introduces a legislative proposal to the EP, the MEPs discuss and propose amendments to the proposal within committees, (some of which are accepted by the committee), and the amended proposal is presented to the plenary session of the EP for the final vote. We curate a dataset of 237,177 amendments consisting of 449,493 edits (changes of contiguous words) across the seventh and eighth terms of EP (2009-2019), and we develop a model to predict which edits are accepted by the committee, based on their text and metadata features and latent representations of the proposing MEPs and laws. The edits proposed by different MEPs often conflict with each other and at most one edit can be accepted from a conflicting set; this naturally leads to the choice of a multinomial logit model for the task. By interpreting the parameters of the learned model, we discover the words, bigrams, and metadata features that are correlated

with edit acceptance, thus giving us valuable insights into the legislative process.

We conclude the thesis in Chapter 6 by summarizing our findings and suggesting directions for future work.

2 Subjective Bias in Documents

In this chapter[1], we propose an interpretable model to score the bias present in web documents, based only on their textual content. Our model is trained on pairs of revisions of the same Wikipedia article, where one version is more biased than the other. Although prior approaches based on absolute bias classification have struggled to obtain a high accuracy for the task, we are able to develop a useful model for scoring bias by learning to perform pairwise comparisons of bias accurately. We show that we can interpret the parameters of the trained model to discover the words most indicative of bias. We also apply our model in five different settings by studying the temporal evolution of bias in Wikipedia articles, comparing news sources based on bias, and scoring bias in law amendments, parliament speeches, and tweets.

In each case, we demonstrate that the outputs of the model can be explained and validated, even for the four domains that are outside the training-data domain. We also use the model to compare the general level of bias between domains, where we see that legal texts are the least biased and news media are the most biased, with Wikipedia articles in between. Given its high performance, simplicity, interpretability, and wide applicability, we expect the model to be useful for a large community, including Wikipedia and news editors, political and social scientists, and the general public.

2.1 Introduction

In recent years, the amount of human-generated data on the Web has increased exponentially in size and relevance. Some prominent examples include encyclopedias (such as Wikipedia), online news portals, and political data (such as legislative acts and speeches). As they are generated by humans, these data suffer from bias to varying degrees, which results from the specific worldview and the motives of their authors.

[1]This chapter is based on (Suresh, Wu, et al., 2023). This author led the project and performed dataset curation, designed the model and experiments for evaluation and application, and interpreted the model and results.

Here we use the term bias to mean *inappropriate subjectivity*, as defined by Pryzant et al. (2020) - "Subjective bias occurs when language that should be neutral and fair is skewed by feeling, opinion, or taste (whether consciously or unconsciously)". Data affected by such subjective bias inform the perspectives and influence the decisions, both political and otherwise, of an increasing number of people. When people are unaware of the bias present in the data, their ability to arrive at accurate conclusions is compromised. This lack of awareness also contributes to the formation of echo chambers and makes it difficult to build consensus for actions for the common good. Therefore, it is important to identify and measure this bias and to do so in an explainable manner so as to be trustworthy and easy to verify.

Currently, this is done manually in several domains: Wikipedia editors mark articles and edits as violating neutrality, companies such as AllSides (AllSides, 2022) provide ratings of bias in the media, and political scientists analyze speeches to study subjective language as expressions of ideological positions. However, such manual analysis cannot scale to the exponentially growing size of web data, hence necessitating the use of automated approaches. Machine-learning models that can benefit from the large training data are of particular interest in this regard.

The English-language Wikipedia is in many ways an ideal source of training data for these models. It has a neutral point of view (NPOV) policy (Wikipedia, 2022c), the adherence to which can be used as a measure of unbiasedness (neutrality). The policy requires following principles such as not stating opinions as facts (and vice versa), not using language that sympathizes with or disparages the subject, etc. Wikipedia also has an active community of editors that enforces this policy by making edits to reword or remove problematic content from articles and leaving comments to indicate NPOV issues. Moreover, the data is extensive due to Wikipedia's vast collection of articles spanning a wide range of subjects; and the complete revision history of these articles, along with the editors' comments, is accessible to the public.

Our goal in this chapter is to develop a model trained on POV-related edits to Wikipedia articles that can quantify bias in web documents and study its applicability to Wikipedia itself, as well as to four domains outside the training data, specifically news, political speeches, legal texts, and tweets. Note that, in addition to being reasonably accurate, we also want the model to be *interpretable*, i.e., we want to use the parameters of the trained model to infer the words indicative of bias and to explain the output of the models. Interpretability is useful for gaining an understanding of how bias occurs and for making the model more trustworthy and easy to debug for users.

Besides enabling us to gain insights into the features indicative of bias, such models could provide a mechanism for quantifying bias in documents across time, topics, and domains. Within Wikipedia, such models can be used to monitor the evolution of bias in articles, as they are revised over time. This can help to draw the attention of editors toward

biased articles and revisions. It would also be useful to study the speed of bias mitigation in peer-produced texts for researchers in this field. News, social media, political speeches, and legal texts are some other domains where bias detection is important, but unlike Wikipedia, data annotated for bias is scarce in these domains, which makes training models difficult. If the models trained on Wikipedia can generalize to these other domains, it can address this problem to some extent.

2.1.1 Absolute versus Relative Bias Classification

Previous work on bias modeling predominantly considers the task of bias classification of short pieces of texts (e.g., words and sentences) in an *absolute* sense, i.e., classifying a given piece of text as biased or unbiased (Z. Li et al., 2022; Pryzant et al., 2020; Zhong et al., 2021). However, we suggest that classifying general web documents in this manner is, for two reasons, not a well-defined task.

First, the threshold for deciding whether a text is biased or not is subjective, especially for longer texts such as documents. In fact, previous work has found poor inter-annotator agreement when obtaining ground-truth labels (De Kock & Vlachos, 2022; Lim et al., 2020; Spinde et al., 2021).

Second, this threshold varies depending on the topic and the domain of the document. For instance, a Wikipedia article considered 'unbiased' on a politically controversial topic is arguably prone to having more subjective statements than a 'biased' one describing an objective scientific truth. A 'biased' Wikipedia article can still be more objective than a relatively 'unbiased' political speech.

Therefore, in this chapter, we instead consider the task of *relative* bias classification of documents. We define this as the task of predicting which text, among a given pair of texts, is more biased. Solving such a task does not require the determination of a bias threshold thus avoids the above two issues. Previous works have found greater inter-annotator agreement and higher human accuracy for this task, compared to absolute-bias classification (Aroyo et al., 2019; De Kock & Vlachos, 2022).

We can obtain abundant training data for this task from the revision history of Wikipedia articles. Each time a Wikipedia editor corrects a POV issue present in an article version, a pair of texts is generated where one text (the version before the correction) is more biased than the other (the version after the correction). Comparing bias at the document level instead of the word or sentence level also enables us to benefit from additional context information such as the overall topic of the document.

Note that, although we train our model for comparing the bias of pairs of texts, it can be applied to score the bias in a single piece of text. This is possible because we use a Bradley-Terry model of pairwise comparisons; it learns a score for items being

compared as discussed in Section 1.4. This score, when parametrized in terms of domain-independent text representations as features, can be interpreted as a *real-valued* measure of bias that can be applied across texts from different topics and domains.

Unlike *binary labels* such as 'biased' and 'unbiased', a real-valued bias score can be assigned without the need for a topic or domain-dependent threshold. Texts from domains/topics prone to greater subjectivity are expected to be assigned a relatively higher bias score *in general* compared to texts from *other* domains/topics, whereas specific texts from these domains/topics will still have a lower bias score than texts with higher bias from the *same* domain/topic. For instance, articles in news media could have a higher bias score in general than Wikipedia articles, but a factual news article can still have a lower bias score than an editorial.

2.1.2 Validity of Wikipedia NPOV

It is worth noting that using Wikipedia's NPOV policy (as interpreted in light of editorial actions to enforce it) as a standard of neutrality is not without criticism. In fact, Matei and Dobrescu (2011) show that NPOV policy itself and its interpretation is subject to much conflict between Wikipedia editors. Keegan and Fiesler (2017) study the evolution of rules in Wikipedia and note that 'rules-in-form', such as the NPOV policy, could still be subject to deliberation and revision for a long period of time.

In a way, this reflects the underlying difficulty of agreeing on a definition of subjectiveness; in other words, subjectivity itself seems to be subjective. This problem is mitigated to some extent if we consider *comparisons* of subjectivity that seem to be more objective based on the fact that there is better inter-annotator agreement as discussed previously. Moreover, as our model is *interpretable*, users can re-assess the factors leading to the assignment of a certain score if needed, so as to reduce the effect of subjectivity in the labels used for training.

An alternate view of the NPOV revision dynamics is provided by Pavalanathan et al. (2018); they observe that the NPOV policy encourages articles to converge to a common standard of neutrality, as judged by several lexicons, in spite of the editors themselves not changing their styles. This suggests that in many cases editors permit edits by others that reduce bias in terms of the community's interpretation of the NPOV policy, even if they themselves would not make such edits or consider them necessary.

We manually analyzed 100 random POV-related edits (identified automatically based on editor comments). We agreed that 82 of them reduced bias, and though we did not think the other 18 necessarily reduced bias, we did not consider them to increase bias or to be significantly harmful to the article. Therefore, overall we think the approach of training a model on POV-related edits to quantify bias (with regard to the Wikipedia community's definition and interpretation of neutrality) is reasonable, given that the

users of the model keep in mind its limitations. We discuss this again, at the end of the chapter under ethical considerations.

2.1.3 Research Questions, Contributions, and Outline

We seek to answer the following research questions:

- **RQ1:** Given a pair of consecutive revisions (versions), of the same Wikipedia article, generated when a POV issue is corrected, how well can we predict which one among them is more biased (i.e., the version before the correction), using only models based on their *textual* content?
- **RQ2:** Can we understand which words are correlated with bias by interpreting the parameters of the predictive models?
- **RQ3:** How widely can such predictive models generalize? Can the models that are trained to compare only consecutive revisions of the same article generalize across time, topic, and domain?
  - **RQ3a:** Can the models compare revisions of the article that are not consecutive and thereby capture the temporal evolution of bias of an article over its entire history?
  - **RQ3b:** Can we compare bias in different Wikipedia articles from different topics by using the predictive models?
  - **RQ3c:** Can the predictive models, which are trained only on Wikipedia, generalize to other domains of text? Can we use them to compare the level of bias between different domains?

We make the following contributions:

- Towards answering RQ1, we develop discrete choice models for relative bias classification that use only the article's textual content as features. We compare the accuracy of our model against both random chance and strong baselines.
- We use the parameters of the trained models to compute a bias score for words that represents the contribution of the word to a document being biased. We use the scores to discover words that are indicative of bias (RQ2).
- We also use the trained models to compute a bias score for documents. Although the models are trained to compare only consecutive revisions of Wikipedia articles, the score that we compute using the models can be used to compare bias in non-consecutive revisions of an article, different Wikipedia articles, and even documents

from different domains including laws and news articles. We analyze the computed scores to answer RQ3.

- We curate new datasets of Wikipedia articles to train and evaluate our models. We release publicly all the datasets and our code.

The chapter is structured as follows. In Section 2.2, we describe the related work. In Section 2.3, we provide details about datasets we use and have curated for this study. In Section 2.4, we describe the bias model in detail. In Section 2.5, we evaluate the performance of the model, explore its interpretability, and comment on some potential applications of the model to other domains. We conclude the chapter in Section 2.6.

2.2 Related Work

Bias in web documents is a well-studied topic. Previous works have studied subjective bias in Wikipedia (De Kock & Vlachos, 2022; Pryzant et al., 2020; Wong et al., 2021; Zhong et al., 2021), news media (Lim et al., 2020) and political speeches (Vafa et al., 2020).

Wong et al. (2021) collect a dataset of pairs of biased and neutral versions of the same Wikipedia article. They do this by going through the revision history of articles and choosing revisions where the POV template (which is used to flag NPOV issues) was added by editors and where it was removed. However, the dataset size is quite small (only about 5,000 pairs) as the template is not updated often. The authors only use metadata to develop classification models to predict if a given article version is biased or not, and they achieve an accuracy only slightly better than random (52%).

Pryzant et al. (2020) and Zhong et al. (2021) work on the task of identifying bias in words and sentences. They obtain a parallel corpus of about 200,000 pairs of biased and neutral sentences by aligning the sentences before and after an NPOV-related revision using the BLEU score. Such revisions are identified by checking for the presence of regular expressions in the editor's comments that denote NPOV issues in the revision. They use this corpus to train models for the binary classification task of predicting whether a word or sentence is biased. For the task of sentence bias prediction, the best model in Zhong et al. (2021) achieves an accuracy of 73%, after additional fine-tuning on a manually annotated set of sentence pairs. Z. Li et al. (2022) consider the problem of bias detection when annotations are scarce, noisy, and potentially biased. By using data augmentation and a self-supervised contrastive learning objective, they are able to achieve a similar performance as Zhong et al. (2021) with a much smaller dataset. None of these works considers the task of predicting bias at the document level.

De Kock and Vlachos (2022) consider the related task of promotional tone detection at the document level. Similarly to Wong et al. (2021), they also use template information

to collect the dataset hence have a relatively small dataset size. They use neural network models with a gradient reversal layer to prevent the model from learning features that are topic specific to enable better generalization. Their best model achieves an accuracy of 64.3% for predicting if a given text has a promotional tone.

The models above are based on deep neural networks hence require significant time and GPU resources for training and inference. In particular for training, the models in Pryzant et al. (2020) and Zhong et al. (2021) need several hours, and the model in De Kock and Vlachos (2022) needs more than a day.

The models are also complex hence difficult to interpret. Although an explanation can be given for which parts of a *given* text are biased, it is difficult to answer, based on the trained model, which words *in general* are indicative of bias.

Moreover, all the models above are trained for the task of bias prediction in an absolute sense, whereas our model is trained for the pairwise comparison of bias, a better-defined task as described in Section 2.1[2].

Our models are instances of the class of discrete-choice models that are regularly used for several applications involving learning from pairwise comparisons. Such models have been used for predicting the survival of edits in Wikipedia, without using the edit text (Yardim et al., 2018), for predicting the outcome of football matches (Maystre et al., 2019), and for predicting the success of amendments in the European Parliament (Kristof et al., 2021). (Kristof et al., 2021) also study the use of text features in combination with other features.

To the best of our knowledge, we are the first to propose modeling bias in documents by using a framework of pairwise comparisons. We use a discrete-choice model based only on the document text. Compared to prior bias models, this model is easily interpretable and relatively inexpensive computationally to train and use, and it also achieves similar or better accuracies. We also study the application of the model in a variety of document domains. We demonstrate its generalizability and describe the insights that could be gathered from it.

2.3 Datasets

We use six datasets in this chapter, three of which we collected ourselves. We present a summary of the statistics of each dataset in Table 2.1. We will now briefly describe the datasets.

[2]De Kock and Vlachos (2022) additionally test their model for ranked prediction of promotional tone and report an accuracy of 74.1% but the model is not trained for this task.

2.3.1 Wikipedia: Article Neutrality

To train and evaluate our model, we curate a new dataset that we call the Wikipedia article neutrality dataset (WAND). The dataset can be viewed as an article-level version of the sentence pair dataset collected in Zhong et al. (2021).

The dataset consists of the text of pairs of revisions of the same Wikipedia article where one revision is more biased than the other. We collect it by going through the revision history of all articles in the English Wikipedia and by collecting a pair of revisions before and after a POV-related edit is made. We identify the POV-related edits by checking for the presence of certain regular expressions; we use the same list of expressions used in Zhong et al. (2021).

For each revision, we use the `mwparserfromhell` package (Kurtovic, 2022) to parse its *wikitext* as obtained from the MediaWiki API (Wikimedia, 2023). We then apply the text pre-processing steps, followed by Wong et al. (2021) and Pryzant et al. (2020), to keep only the plain text (excluding wikilinks, templates, and tags) from the main content part of the article (excluding the External Links and References sections).

For every pair, we compute the Levenshtein edit distance between the texts of the revisions and keep only the pairs that have a distance of at least ten. We do this to remove pairs where only minor edits such as corrections in spelling and punctuation were made.

2.3.2 Wikipedia: Controversial Issues

As the WAND dataset contains the revisions at only the times of the POV-related edits, we cannot use it to evaluate the performance of our models in estimating bias changes over the whole history of the articles. Therefore, we construct a new dataset of revisions of the articles mentioned in Wikipedia's *List of Controversial Issues* (Wikipedia, 2022a). Wikipedia editors are urged to regularly check these articles to make sure that the presentation follows the NPOV policy, as they are frequently subjected to biased edits. The articles in the list are organized by topics, such as *Politics and Economics*, *History*, *Religion*, etc.

For each article, we collect the text for 100 revisions periodically sampled from its history. The number of revisions k between each sampled revision is different for each article, as some articles have more revisions in their histories than others (depending on age or frequency of editing of the article). The text is pre-processed, as in WAND, to retain only the plain text from the main article content.

2.3.3 News

To study how well models of bias generalize to domains that are different from their training domain, we apply our models trained on Wikipedia to score bias in news articles. We use the Webis Bias Flipper-18 dataset (Chen et al., 2018) that contains news articles from outlets with different ideological biases (left, right, and center). The articles that describe the same event are grouped into stories, which enables us to eliminate the effect of the event itself by ranking articles within each group.

The grouping of the news articles and the ideological bias labels of the outlets come from AllSides.com (AllSides, 2022). This website aims to present balanced coverage of news by presenting articles from outlets with different ideological biases. The ideological bias labels for each outlet are determined by a combination of factors, including editorial review and community feedback.

2.3.4 European Parliament: Law Amendments

Texts in the legal domain are likely to have much less subjective language relative to general texts. To see if the model can score these correctly, we evaluate it on a dataset of law amendments. We use the dataset of amendments proposed in the eighth term of the European Parliament, released by Kristof et al. (2021).

Each amendment in the dataset consists of a pair of texts. The first text is a paragraph of the law in the original law text, as drafted by the European Commission, the executive branch of the European Union and the body in charge of drafting new laws. The second text is the amended version of the same paragraph as proposed by a parliamentarian or a group of parliamentarians when the law is being discussed within the committees of the European Parliament. An amendment consists of multiple edits, where an edit is a contiguous block of text that is deleted, inserted, or replaced.

Each proposed amendment is voted on within the committee, and only a subset of its edits may be accepted for incorporation into a modified draft law that is subsequently presented at a plenary meeting of the parliament.

2.3.5 European Parliament: Debates

We also study the ability of our model to generalize to the domain of political speeches. For this, we use a dataset of debates scraped from the European Parliament website.

A debate consists of multiple speeches where the same law or resolution is discussed. For each speech, we know the speaker and the European political party to which they belong. We obtain the political family of each party (Center, Radical Left, Conservative, etc.) from the Chapel Hill Expert Survey (CHES) data (Jolly et al., 2022).

2.3.6 Climate Change Tweets

Finally, we evaluate the model on text from Twitter. We use the Twitter Climate Change Sentiment Dataset on Kaggle (Qian, 2019) for this purpose. The dataset contains tweets pertaining to climate change, collected between April 27, 2015, and February 21, 2018. Each collected tweet was independently annotated by three reviewers into one of three categories:

- *Anti*: Tweet rebuts belief in man-made climate change
- *Pro*: Tweet supports belief in man-made climate change
- *News*: Tweet links to factual news about climate change

Only those tweets where all three reviewers agreed have been included in the dataset.

2.4 Model

We now describe the model we propose for ranking documents by bias. Model interpretability is one of our primary concerns. Hence, taking this into account, we design the whole pipeline from feature extraction to prediction. In particular, we generally avoid using multi-layer neural networks, except in some models where we use it for feature extraction at the word level.

2.4.1 Features

To represent the text of a document, we use the normalized sum of the embedding vectors of the words in the text. We experiment with both static and contextual word embeddings.

Static word embeddings represent the meaning of each word by a single vector in d-dimensional space. If a word has different meanings in different contexts, the single vector represents a weighted average of the different meanings based on their frequency. The embedding is obtained by training each word's embedding vector so as to predict the words that appear in its context, given their embedding vectors. We use the pre-trained fastText (Unsupervised) embeddings (Bojanowski et al., 2017c) that were trained on the English Wikipedia. More details regarding static word embeddings are given in Section 1.2.2.

Contextual word embeddings, on the contrary, represent the meaning of a word in the context where it appears, hence each *occurrence* of the word (a *token*) is represented by a single vector. The BERT model (Devlin et al., 2019) is arguably one of the most

Wikipedia Article Neutrality	
Number of articles	358,941
Number of revision pairs	895,957
Median revision pairs per article	1

Wikipedia Controversial Issues	
Number of articles	1,544
Median history length	4,729

News	
Number of stories	2,781
Number of outlets	77
Number of articles	6,448

European Parliament Amendments	
Number of original texts	28,407
Number of proposed amendments	98,245
Amendments with at least 1 accepted edit	37,689
Amendments with at least 1 rejected edit	73,604

European Parliament Debates	
Number of debates	3,404
Number of political party families	10
Number of speeches	104,651

Climate Change Tweets	
Number of Anti tweets	3,990
Number of Pro tweets	22,962
Number of News tweets	9,276

Table 2.1: Dataset statistics

commonly used contextual embeddings and has been used in prior work in bias modeling at the sentence level (Zhong et al., 2021). However, it can model sequences only up to a maximum length of 512 tokens due to the quadratic complexity of the attention mechanism, hence cannot effectively model long documents such as Wikipedia articles.

Therefore, we use a pre-trained Longformer model (Beltagy et al., 2020), which is a variation of BERT that uses sliding window attention, thus enabling it to model long sequences efficiently. Specifically, we use the `longformer-base-4096 model` from HuggingFace (Allen Institute for AI, 2022). It has been used to model Wikipedia articles in prior work (De Kock & Vlachos, 2022). More details regarding contextual word embeddings are given in Section 1.2.4.

We convert all text to lowercase before computing the vector representation. When using static embeddings, we obtain the vector representation of a text i as

$$\widehat{\mathbf{t_i}} = \frac{\mathbf{t_i}}{\|\mathbf{t_i}\|}, \quad \widehat{\mathbf{t_i}} \in \mathbb{S}_1^d \text{ (unit sphere)}, \tag{2.1}$$

where

$$\mathbf{t_i} = \sum_{w \in \mathcal{V}_i} n_i(w) \mathbf{v_w}. \tag{2.2}$$

Here $\mathcal{V}_i$ is the set of words in text i, $n_i(w)$ is the frequency of word w in text i and $\mathbf{v_w}$ is the embedding vector of the word. The representation $\mathbf{t_i}$ is obtained in a similar manner when using contextual embeddings except that we consider tokens instead of words.

2.4.2 Model Architecture

Our model takes inputs in the form of *pairs* of texts and predicts which text is more biased than the other. We use the Bradley-Terry model of pairwise comparison outcomes (Bradley & Terry, 1952).

We define the probability that text i is more biased than text j to be

$$P(i \succ j) = \frac{e^{s_i}}{e^{s_i} + e^{s_j}}, \tag{2.3}$$

where $s_i, s_j \in \mathbb{R}$ are *bias scores* of texts i and j, respectively (higher means more biased).

To simplify the rest of the description, we assume that we are using static word embeddings. It is straightforward to extend this to the case of contextual word embeddings, by using tokens in place of words.

We model the bias score of a text i as the sum of the bias contributions of the words present in the text, weighted by the number of times each word occurs in the text. More precisely, we define

$$s_i = \frac{1}{K_i} \sum_{w \in \mathcal{V}_i} n_i(w) B(w, i), \tag{2.4}$$

where $B(w, i)$ is the bias contribution of the word w given the topic of text i. We also include a scaling factor K_i to ensure that the bias score of a text does not depend on its length or generality. This enables us to compare the bias within a diverse set of texts. We explicitly define K_i later in this section.

We model the bias contribution $B(w, i)$ as a function of both the word w and the text i, as the bias induced by words can change depending on the topic of the text. For instance, the word *malicious*, when used as an adjective to describe the nature of a specific person, usually indicates bias, but when used within a computer science article, it can be legitimate (e.g.,*malicious code*).

To model this we define $B(w, i)$ as

$$B(w, i) = \mathbf{f_i}^T \mathbf{v_w}, \tag{2.5}$$

where $\mathbf{f_i} \in \mathbb{R}^d$ is the bias word *query vector* for text i and $\mathbf{v_w} \in \mathbb{R}^d$ is the embedding vector of word w. The smaller the angle between $\mathbf{f_i}$ and $\mathbf{v_w}$ is, the higher the bias contribution of w given the topic of text i.

The query vector $\mathbf{f_i}$ depends on the topic of text i. We model it as an affine function of the vector representation $\widehat{\mathbf{t_i}}$ of the text i,

$$\mathbf{f_i} = \mathbf{W}^T \widehat{\mathbf{t_i}} + \mathbf{b}, \tag{2.6}$$

where $\mathbf{W} \in \mathbb{R}^{d \times d}$ and $\mathbf{b} \in \mathbb{R}^d$ are learned parameters. This simple formulation enables us to easily compute a general (topic-independent) version of the word bias score that we describe later.

Substituting (2.6) in (2.5), and (2.5) in (2.4), and using (2.1) and (2.2) to simplify, we get the bias score of the text as

$$s_i = \frac{\|\mathbf{t_i}\|(\widehat{\mathbf{t_i}}^T \mathbf{W} \widehat{\mathbf{t_i}} + \mathbf{b}^T \widehat{\mathbf{t_i}})}{K_i}. \tag{2.7}$$

We can see from (2.2) that the quantity $\|\mathbf{t_i}\|$ depends on the total number of words in the text. If a text is concatenated with itself, $\|\mathbf{t_i}\|$ will increase even though the content and bias of the text do not change.

Also, if two texts i and j are similar (i.e., $\widehat{\mathbf{t_i}}$ and $\widehat{\mathbf{t_j}}$ have high similarity) and therefore should have similar bias, but i is more specific and uses a less diverse set of words than j (i.e., the embeddings $\mathbf{v_w}, \forall w \in \mathcal{V}_i$ have a lower variance than the embeddings $\mathbf{v_x}, \forall x \in \mathcal{V}_j$), then $\|\mathbf{t_i}\|$ tends to be larger than $\|\mathbf{t_j}\|$. This could happen for instance if j gives some context around the topic, placing it within a more general topic.

Since we would like the bias score of the text to not change in these cases, we define the scaling factor to be $K_i = \|\mathbf{t_i}\|$. We then have

$$s_i = \widehat{\mathbf{t_i}}^T \mathbf{W} \widehat{\mathbf{t_i}} + \mathbf{b}^T \widehat{\mathbf{t_i}}. \tag{2.8}$$

To interpret the model to identify the bias of words, we need to get the true values of all $B(w, i)$, for which we need precise inference to be possible for $\mathbf{W}$ and $\mathbf{b}$ (i.e., the model should be identifiable). It is straightforward to see that this is satisfied if and only if $\mathbf{W}$ is symmetric[3]. We therefore parameterize $\mathbf{W}$ as

$$\mathbf{W} = \mathbf{U} + \mathbf{U}^T, \tag{2.9}$$

where $\mathbf{U} \in \mathbb{R}^{d \times d}$ is the variable that is optimized during learning.

While $B(w, i)$ gives the bias contribution of word w when it appears in text i, we are also interested in obtaining the general bias score of a word in a given corpus of texts $\mathcal{C}$ without specifying any particular text. Hence we define the general bias score of a word w as an average of its bias score over all texts, i.e.,

$$GB(w) = \frac{\sum_{i \in \mathcal{C}} B(w, i)}{|\mathcal{C}|} = \bar{\mathbf{t}}^T \mathbf{W} \mathbf{v_w} + \mathbf{b}^T \mathbf{v_w}, \tag{2.10}$$

where

[3]This follows from the fact that $\widehat{\mathbf{t_i}}^T \mathbf{W} \widehat{\mathbf{t_i}}$ is a quadratic form, and that over real numbers there is a one-to-one correspondence between such quadratic forms and symmetric matrices $\mathbf{W}$ that determine them.

$$\bar{\mathbf{t}} = \frac{\sum_{i \in \mathcal{C}} \mathbf{t_i}}{|\mathcal{C}|}. \tag{2.11}$$

Note that the affine formulation of $\mathbf{f_i}$ enables us to compute $GB(w)$ by averaging the text representations $\mathbf{t_i}$ separately, thereby reducing the computational complexity.

$GB(w)$ can be extended to the case of contextual word embeddings by averaging the bias score over each occurrence of a word. However, in this chapter, we restrict it to models using static embeddings as it is significantly easier to compute in that case.

We call a version of our model including only the linear term $\mathbf{b}$ in (2.6) as the *Linear* model and the full model including both terms as the *Quadratic* model.

2.4.3 Training

We use the WAND dataset for training. We split the revision pairs into training, validation, and test sets in the ratio 90:5:5. To avoid data leakage, we take care to ensure that all pairs from a given article are present in the same split.

We train each model by maximizing the likelihood of the training data, under the probability model in (2.3). More precisely, we solve the optimization problem given by

$$\max_{\theta} \prod_{(i,j) \in \mathcal{D}} P(i \succ j | \theta), \tag{2.12}$$

where $\theta = \{\mathbf{U}, \mathbf{b}\}$ is the set of parameters to be learned, $(i, j) \in \mathcal{D}$ are the revision pairs in the train set (i is the version before the edit, j is the version after the edit), and $P(i \succ j | \theta)$ is the probability that i is more biased than j given the parameters θ, modelled as in (2.3).

We use mini-batch stochastic gradient ascent for the maximization. Models based on static embeddings take approximately 2 hours to train, while those using contextual embeddings take approximately 2 days to train on an NVIDIA Quadro RTX 6000 GPU. For the models using the pre-trained contextual embeddings, we keep the weights of the embedding model fixed during training (i.e. no fine-tuning) due to constraints on GPU usage time. We do not observe any overfitting based on the performance of the model on the validation set and therefore do not use any regularization.

2.5 Evaluation and Applications

In this section, we evaluate the performance of our models, examine their interpretability and explore their applications in a variety of domains.

2.5.1 Evaluation

We evaluate the generalization ability of our models on the task of pairwise bias classification by measuring their accuracy on the test set. Since we are comparing the bias within a pair and a difference is always present within a pair (both versions do not have exactly the same bias), the concept of type I and type II errors are not as relevant and accuracy is a sufficiently informative metric.

We compare against several baselines which we describe below:

- **Random**: The random classifier predicts one of the two versions in a pair uniformly at random to be the more biased one.
- **Wiki Words to Watch (Words2Watch)**: Wikipedia maintains a list of words that could potentially cause bias called *Words to Watch* (Wikipedia, 2022b). This classifier first compares the count of such words among the two versions in the pair. If the counts are equal, it picks one of two versions uniformly at random. Otherwise, it predicts the version having the higher count to be the more biased version.
- **Sentence Bias Aggregate (SentAgg)** These classifiers are based on the sentence-level bias models developed by Zhong et al. (2021).

  We first extract a dataset of biased and neutral sentences from the revision pairs in our train set, following the procedure in Pryzant et al. (2020). Like Zhong et al. (2021), we then train sentence-level classifiers based on the pre-trained BERT model on this dataset to predict if a sentence is biased or neutral.

  To perform the version level bias comparison for a pair, SentAgg aggregates the predicted probability for the sentences in each version and compares the aggregated probability of the two versions.

  We construct three versions of the SentAgg classifier:

  - *SentAgg-Max*, where BERT weights are not finetuned when the sentence classifier is trained and aggregated probability for a version is computed as the *maximum* of the predicted probabilities for the sentences in the version,
  - *SentAgg-Mean*, where BERT weights are not finetuned but the aggregated probability is computed as the *mean* of the predicted probabilities and

Model	Accuracy(%)
Random	50 ± 0.46
Words2Watch	63.4 ± 0.44
SentAgg-Max	55.33 ± 0.46
SentAgg-Mean	68.35 ± 0.43
SentAgg-FT-Mean	76.01 ± 0.39
Static Linear	75.29 ± 0.40
Static Quadratic	$\mathbf{76.84 \pm 0.39}$
Contextual Linear	74.35 ± 0.40
Contextual Quadratic	$\mathbf{77.56 \pm 0.38}$
Human	74.00 ± 8.60

Table 2.2: Accuracy of models

– *SentAgg-FT-Mean*, where BERT weights are *finetuned* when the sentence classifier is trained and aggregated probability is computed as the *mean* of the predicted probabilities.

The test accuracy of all baseline models and our models, and their 95% confidence intervals are given in Table 2.2. The accuracies of the top two models are highlighted in bold. We also include a human performance benchmark which was obtained by one of the authors manually labeling 100 randomly chosen pairs from the test set.

The best performance of 77.56% accuracy is achieved by our *Quadratic* model using contextual word embeddings, and the same model using static word embeddings achieves a close second with an accuracy of 76.84%. Both models significantly exceed the performance of all baselines. The higher accuracy achieved by our *Quadratic* models relative to our *Linear* models suggests that the information given by the document topic in computing $B(w, i)$ is beneficial.

The best-performing baseline is *SentAgg-FT-Mean* which uses a fine-tuned BERT model. Note that our models achieve a better performance than it does despite not using fine-tuning. The large difference in performance between *SentAgg-Mean* and *SentAgg-FT-Mean* suggests that fine-tuning could further improve the performance of our models to some extent, albeit with much higher computational costs.

We now compare the models in terms of their inference time, i.e., the time taken for the trained model to score and compare a pair of versions.

Remarkably, our *Static Quadratic* model outperforms *SentAgg-FT-Mean* and achieves a performance similar to *Contextual Quadratic* while requiring much less computational resources than both those models for training and inference. From Table 2.3, we see that inference in *Static Quadratic* is almost an order of magnitude faster than *SentAgg-FT-*

Model	Inference time(ms)
Static Quadratic (on CPU)	130
Contextual Quadratic	816
SentAgg-FT-Mean	1,278

Table 2.3: Inference time of the best-performing models per version pair, averaged over 1,000 randomly chosen pairs from the test set.

High $GB(w)$			**Low** $GB(w)$
1-10	**11-20**	**21-30**	$\mathbf{P_{10\%}}$
impressive	stunning	spectacular	waived
finest	horrible	arrogant	readings
superb	splendid	memorable	discussed
wonderful	talented	awesome	convened
toughest	amazing	magnificent	attended
formidable	pleasing	ruthless	supplements
brilliant	proud	daring	chaired
exciting	fascinating	greatest	grams
beautiful	clever	courageous	served
excellent	terrible	incredible	suggested

Table 2.4: Words w in decreasing order of $GB(w)$

Mean, while also not using a GPU.

In addition to being fast and accurate, the *Static Quadratic* model is also highly interpretable as it doesn't use deep neural networks. In the experiments that follow where we illustrate our model's interpretability and its application in diverse domains, we use the *Static Quadratic* model unless mentioned otherwise.

2.5.2 Interpretation

A salient feature of the model is its ability to provide explanations for its bias scoring, by computing scores for individual words in the text. We interpret the trained model to see the words indicative of bias. First, we obtain the general bias score $GB(w)$ for every word w in the WAND dataset. The list of top 30 words with the highest $GB(w)$, and the list of 10 words at the 10^{th} percentile are given in Table 2.4.

We see that the words with the highest scores are typically subjective adjectives and other subjective words. The words with lower scores are typically verbs and common nouns.

To have a more comprehensive analysis, we plot in Figure 2.1 the part-of-speech (POS) distribution of the top 1,000 words in terms of $GB(w)$ in comparison to that of all

Word Type	**Mean** $GB(w)$
All	48.84 ± 0.28
Words2Watch	108.21 ± 14.30

Table 2.5: Mean $GB(w)$ of all words vs Words2Watch

words. We see clearly that the proportions of adjectives (ADJ) and adverbs (ADV) in the bias-inducing words are significantly higher than that of all words, while the proportion of proper nouns (PROPN) and common nouns (NOUN) are significantly lower. The proportion of verbs (VERB) is nearly the same.

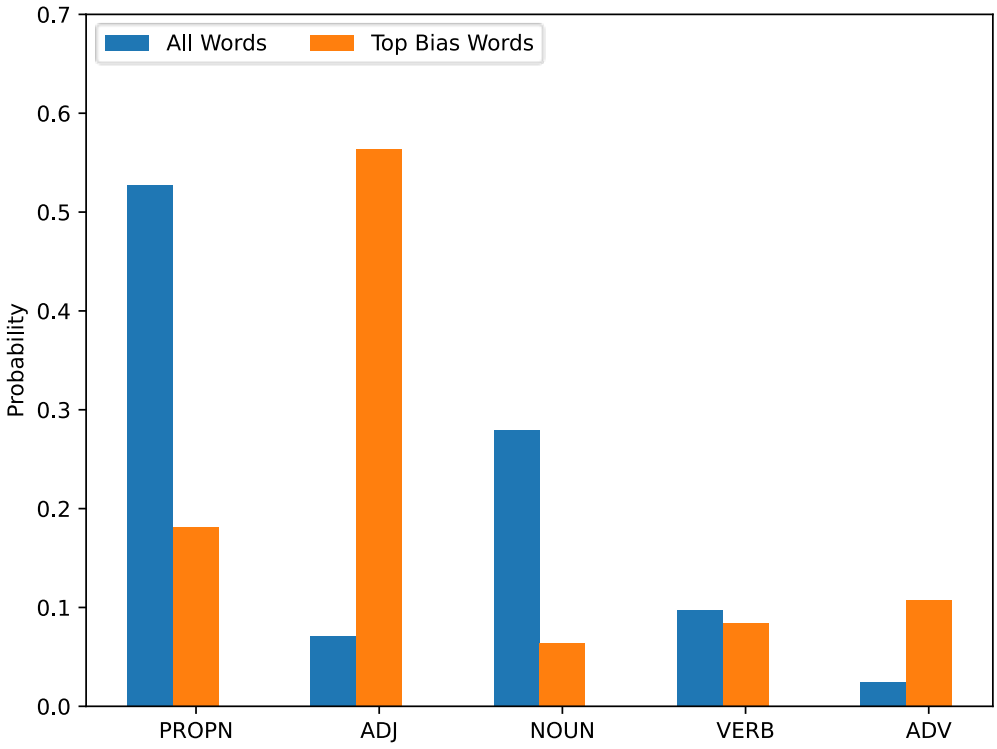

Figure 2.1: Comparison of POS distributions

To provide external validation for the word bias scores $GB(w)$ generated by the model we rely on the Wikipedia *Words to Watch* list. In Table 2.5, we give the mean $GB(w)$ of all words as well as the words in the *Words to Watch* list, with their 95% confidence intervals. We see clearly that mean $GB(w)$ of *Words to Watch* is significantly higher than that of all words.

We can also compare the values of $B(w, i)$ for the same word in different articles to see how the bias induced by the word changes depending on the article's topic. For instance, the word *poorly* when used in the sense of bad performance in sports (in the article *Howard Johnson* (baseball player)) has a $B(w, i)$ score of 324.12. In contrast, it has a

Another change was that apart from no drummer appearing on the album all guitars were recorded directly into the mixing desk without a guitar amp. This is **without a doubt the most** brutal album **ever** made without a drumkit and guitar amp. The spontaneity brought the focus away from feats of musicianship and sent it towards monstrous sounding riffs and **great** songs.

Table 2.6: An excerpt from the article *The Berzerker*, a death metal band. Words with the highest bias according to the *Contextual Quadratic* model are highlighted in bold. The highest bias words according to the *Static Quadratic* model are underlined.

much lower bias score of 30.56 when used to describe something 'burning poorly' in the article *Hydrogen Storage*.

The *Contextual Quadratic* model can also be interpreted to identify words and especially multi-word phrases that induce bias. An example is shown in Table 2.6, where the model correctly identifies the bias-inducing phrase *without a doubt*, which is also mentioned as part of Wikipedia's *Words to Watch*. The *Static Quadratic* model fails to identify the phrase and incorrectly identifies *amp* to be a bias word.

We now comment on some applications of our model for scoring bias in different settings. Note that the model has only been trained on Wikipedia data. In each case, we apply the same preprocessing steps to the text as we did while training.

2.5.3 Bias in Wikipedia

We first apply the model to its training domain, namely scoring the bias in Wikipedia articles. We use the data in Wikipedia: Controversial Issues dataset for the analysis in this section.

Article-level bias

First, we compute the average bias score of each article across its revisions and identify the articles with the highest and lowest scores. The results for the articles within the *Politics and Economics* section of the dataset are given in Table 2.7. We see that the articles with the highest scores are about subjective topics like different '-ism's, and highly controversial topics like racism and denial of genocide. By comparison, the articles with the lowest scores tend to be about fairly objective topics (although still controversial, as we are comparing within the list of controversial topics) like Macedonia, CBC News, and the National Rifle Association. The article on Russian interference in US elections, although it deals with a controversial topic, is well-sourced and protected.

Highest mean s_i	Lowest mean s_i
Anti-Italianism	Macedonia
Patriotism	National Rifle Association
Anti-Irish racism	CBC News
Genocide denial	Federal Marriage Amendment
Black Supremacy	Russian Interference...

Table 2.7: Most and least biased articles in the *Politics and Economics* section

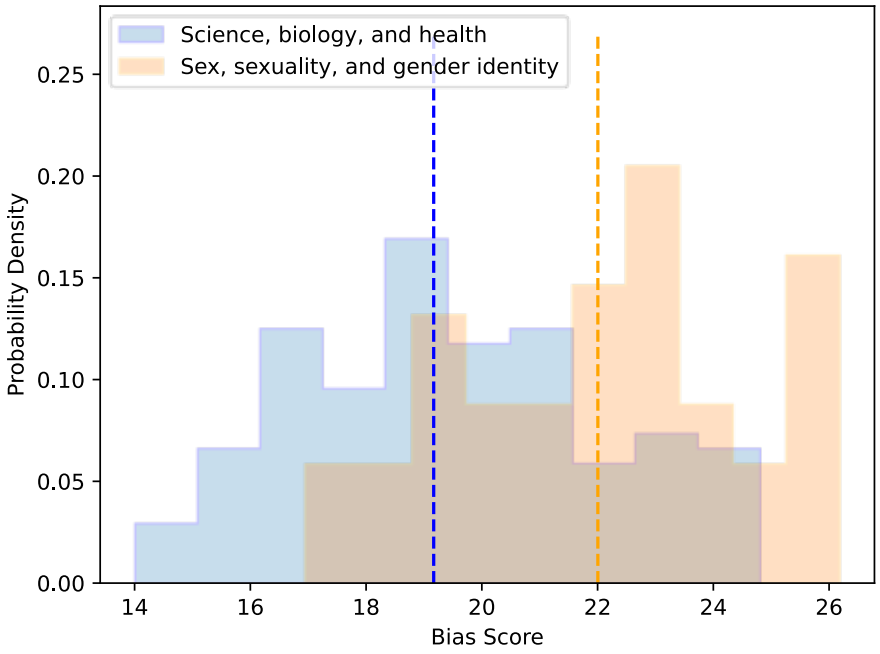

Figure 2.2: Distribution of bias scores within topics.

Topic-level bias

Second, we compare the distributions of bias scores of articles in two different topics, namely *Science, biology, and health*, a relatively objective topic, and *Sex, sexuality, and gender identity* which contains articles on highly controversial topics such as gay rights. The distributions are given in Figure 2.2. The vertical bars show the positions of the means.

We see that the articles in the *Sexuality* topic generally have a higher bias score, as expected. There is some overlap as many articles such as *Abortion*, *AIDS*, etc. occur in both topics.

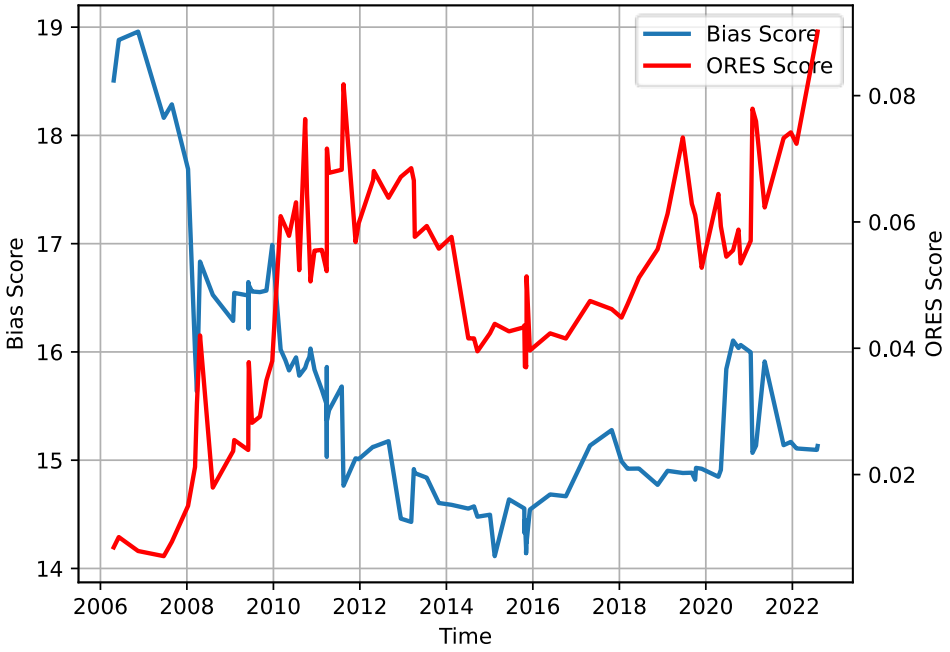

Figure 2.3: Bias score of *Heritability of IQ* over time. Wikipedia ORES article quality scores are plotted for comparison. Spearman correlation: -0.27, p-value 0.008.

Temporal Evolution of Bias

Third, we study the evolution of bias over time by plotting the bias score s_i over time as revisions are made to an article. As examples, we show the plot of the article *Heritability of IQ* in Figure 2.3.

For comparison, we also plot in the same figure the article quality score computed by the Wikipedia Objective Revision Evaluation Service (ORES) (Halfaker & Geiger, 2020). ORES uses a machine learning model to predict article and edit quality based primarily on structural features (templates, headings, links, citations, etc.) and some textual features (length, Words to Watch matches, etc.). Wikipedia editors use the ORES article quality scores to periodically evaluate articles and identify the ones to focus on for editing (such as popular articles that have low quality).

We refer to the probability that ORES gives for an article to be a 'Good Article' as per Wikipedia's criteria (Wikipedia, 2023b) as the *ORES score*. Importantly, one of the criteria is that the article should be neutral as per Wikipedia's NPOV policy. The ORES score for a revision can be obtained by querying a public API (Wikipedia, 2023a).

We see from Figure 2.3 that the bias score computed by our model has a negative

correlation with the ORES score, which is expected as bias negatively affects quality. The median Spearman correlation across all articles in the dataset is -0.27 and the example in Figure 2.3 has the same correlation.

Note that we do not expect a very high magnitude of correlation since ORES also considers other aspects of article quality besides bias and primarily uses structural features which our model does not use. The advantage of our model is that it does a much better assessment of the bias of the textual part, as it is not restricted to a manually built lexicon of bias words unlike ORES (cf. the performance of Words2Watch baseline in Table 2.2). Also, the fact that our model only uses textual features enables it to generalize to other domains as we show in subsequent sections.

We manually examined the text of the revisions of *Heritability of IQ* at different points in time to examine the reason behind the bias trend seen in Figure 2.3. We see that the bias is relatively high in the beginning, because the article has been written primarily by a single author and lacks a balanced view. Over time other authors incorporate competing views and rephrase statements to reflect the presence of controversy. The wording is also improved to make it sound more objective without changing the meaning (eg: *stated* instead of *claimed*). This causes the bias score to generally decrease over time. However, there are instances where an editor inserts or deletes a large amount of biased text that causes sharp fluctuations in the bias score as seen in Figure 2.3. In most cases though, this is rapidly reversed by other editors. We believe this behavior is generally followed in the case of other articles as well, based on a quick perusal of their revision histories.

In addition to the evolution of bias for individual articles, we also study the trend of the average bias across articles over time. This would help to answer questions such as whether on average the bias of an article decreases over time in Wikipedia (RQ3), and if so how fast it decreases.

We consider all articles in the dataset that were created around the same time (in 2003 or 2004), and average each of their bias scores at the same points in time throughout their history. We get the trend shown in Figure 2.4, where the dark line is the average bias score and the shaded area indicates the 95% confidence interval.

We can clearly see that on average the bias of an article decreases over time until it reaches a steady state and that it reaches this state in about ten years. The increasing trend of the ORES score also supports this conclusion.

2.5.4 Media Bias

We now apply the model to score bias in the domain of news media, a different domain from its training domain of Wikipedia. We use the News Dataset in this analysis and we treat the text content of each news article as a document.

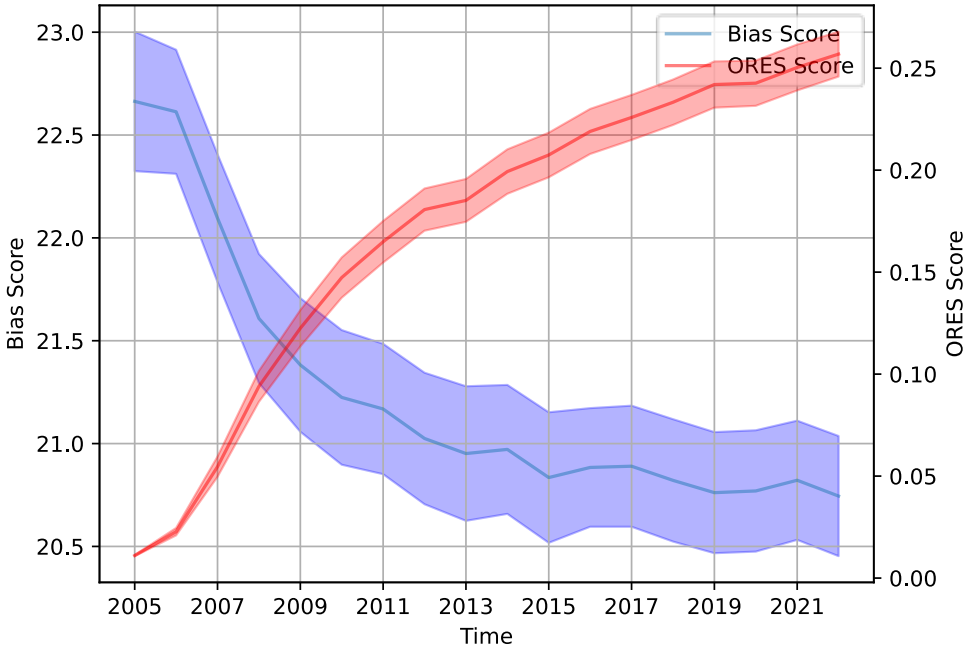

Figure 2.4: Average bias score and ORES score of articles over time.

First, we try to estimate the relative bias level of different outlets to see if we can rank them and identify the ones that are most and least biased. We note that some of the outlets in the dataset, despite having clearly recognized political leanings, also publish many articles verbatim from news agencies such as Reuters and Associated Press. Including such articles would unfairly lower the bias score for these outlets because the articles from news agencies are known to use objective language. We remove such articles from the analysis (except for the articles published by the news agencies themselves). We then obtain a bias score for the remaining articles using our trained model.

We are primarily interested in the relative bias of left and right outlets compared to the center. Hence we include only the stories that have at least one article from a center outlet. Then for every news story, we order the articles covering the story in terms of the bias score and compute the percentile bias score for each article in the story. Finally, we compute the average of the percentile bias scores of the articles from a news outlet to get the mean percentile bias score of the outlet.

We plot the mean percentile bias scores of the outlets along with their 95% confidence intervals in Figure 2.5. For clarity, we only show in the plot the 6 outlets with the smallest confidence interval from each category (left, right, and center). We also show the confidence intervals of the mean percentile bias scores of left, right, and center articles, including those from outlets not shown in the figure.

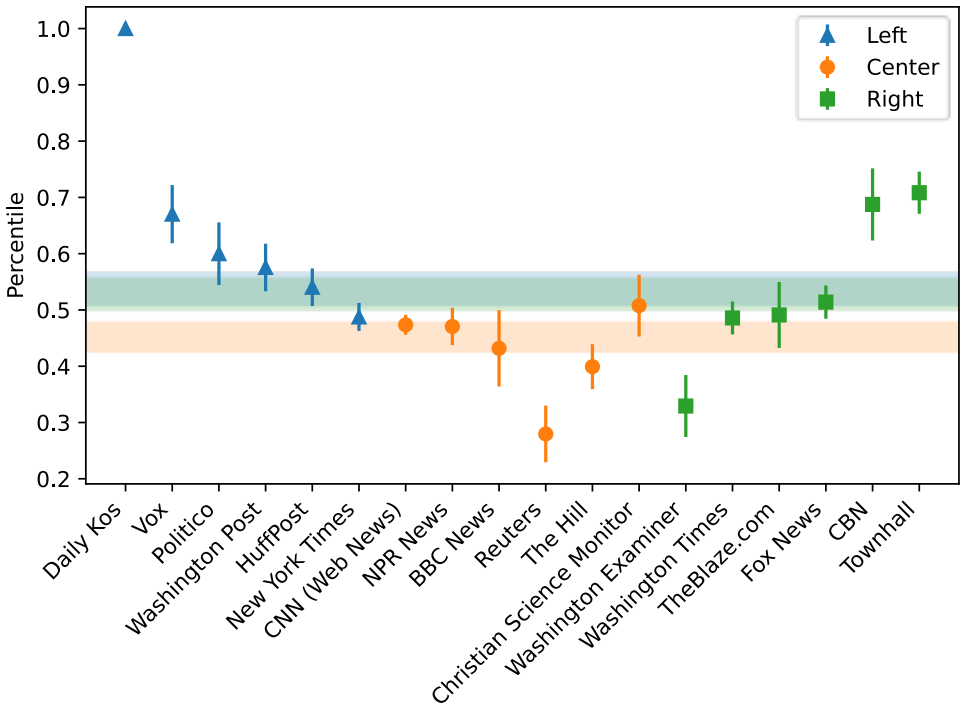

Figure 2.5: Mean percentile bias score of news outlets. The bands show the confidence interval of the mean percentile bias score for left, right and center articles. These include articles from outlets not shown in the figure.

Although there is overlap between individual outlet scores, we see from the confidence intervals of the mean scores that articles from center outlets have significantly lower mean score than those from left and right outlets. Looking at the individual outlet scores, we see that the *Reuters* news agency which is known for its policy of objective language has the lowest bias score. *The Hill*, which claims to provide "objective" and "non-partisan coverage", also has a relatively low bias score. On the other hand, outlets like *Daily Kos* on the liberal side and *Townhall* on the conservative side are open about their political bias. Their articles commonly include partisan commentary on news events and consequently have a very high bias score. This is also true for outlets such as *Vox* and *Christian Broadcasting Network (CBN)* that are relatively less open about their bias. Mainstream outlets such as *New York Times* and *CNN* have a similar and moderate level of bias. *Washington Examiner* is an outlier; it is considered by AllSides to have a Lean-Right bias but has a quite low mean bias score. On manually examining their articles in our dataset, we find that bias appears here in the form of giving a greater fraction of coverage to certain views, rather than word choice or other forms of subjective language. Our model is not expected to detect such forms of bias which explains the low bias score.

Finally, we plot the distribution of bias scores of all the news articles in Figure 2.6, along

Legal text	Mean Score
All (Original + Amendments)	11.34 ± 0.04
Original	9.70 ± 0.10
Amendments	11.81 ± 0.05
Amendments with at least 1 edit accepted	11.64 ± 0.08
Amendments with at least 1 edit rejected	12.01 ± 0.06

Table 2.8: Mean bias score of legal texts.

with the distribution of scores in Wikipedia. We see that the scores are generally higher, as news articles frequently contain subjective commentary on events as discussed earlier, while this is disallowed in Wikipedia.

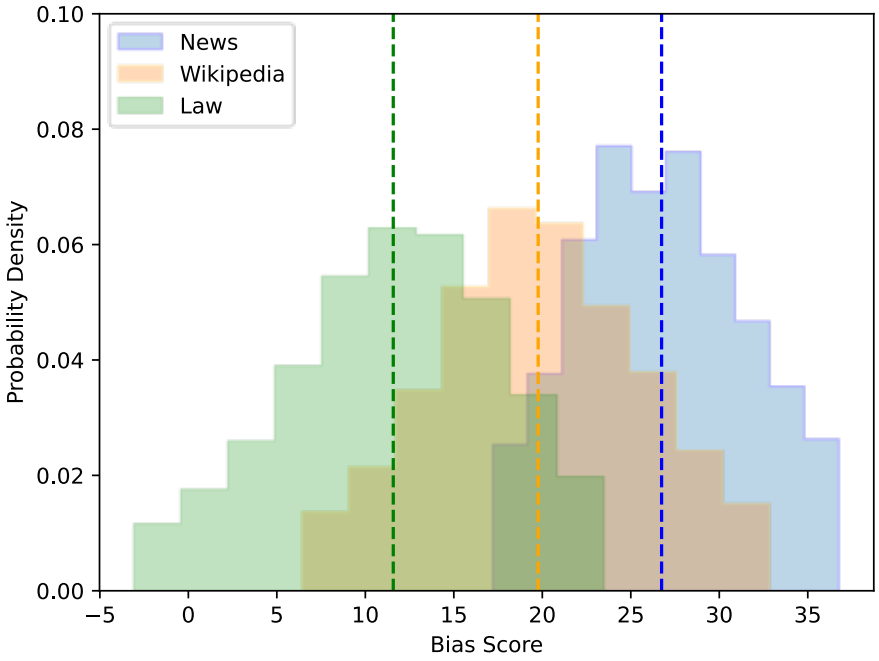

Figure 2.6: Distribution of bias scores across domains

2.5.5 Bias in Legal Texts

We use the European Parliament Amendments dataset to study bias scoring in the legal domain. We give in Table 2.8 the mean bias scores of the different subsets of legal texts in the dataset.

First, we see that the magnitude of bias scores is significantly lower than that of Wikipedia, as is also clear from the distribution of bias scores in Figure 2.6. This is the opposite of

member states ~~may benefit from~~ *have not benefitted from* support in addressing challenges as regards the design and implementation of structural reforms . ~~these challenges may be dependent on various factors , including limited administrative and~~ **institutional** ~~capacity or inadequate application and implementation of union legislation .~~ *on the contrary, the reforms have massively swelled the ranks of the unemployed and helped social inequalities to take hold within the eu , while leaving increasing numbers of citizens* ***destitute*** *by the day .*

Table 2.9: A proposed amendment to the law on *Establishment of the Structural Reform Support Programme for the period 2017 to 2020*. The most biased word in the original text and added text (italicized), is highlighted in bold. The bias score increased from -0.41 to 19.61

what was observed in the case of News. This could be due to the fact that legal texts are carefully crafted to be objective so as to minimize ambiguity in the interpretation of the law. They also tend to avoid partisan language, especially in the introduction sections, so that the text can be accepted by a broad set of legislators who may have diverse viewpoints.

Interestingly, we see that the average bias of the amendments that the parliamentarians propose is higher than that of the original text proposed by the commission. On manually examining the amendments with the highest difference in bias scores, we see that many of them change the introductory sections of the law (explanatory memoranda, recitals etc.) by introducing partisan and subjective language. An example is given in Table 2.9. Nevertheless, we see that among the proposed amendments, the ones that get accepted have relatively a smaller bias on average.

2.5.6 Bias in Political Speeches

We now apply the model to the domain of political speeches. We use the European Parliament Debates dataset for this analysis.

We perform a similar procedure as for News. We consider the text of each speech as a document and obtain the bias score for it. Then we order the speeches in the debate by their bias score, and compute the percentile bias score for each speech. Finally, we compute the average of the percentile bias score of the speeches from the parties belonging to the same family to get the mean percentile bias score of the family.

We show the mean percentile bias scores of all the families in Table 2.10.

We see that as expected, *Center* family has the lowest mean percentile bias score among all families, while the *Green* and *Confessional* (religion-based politics) families have high

Family	Mean Percentile
Agrarian/Center	40.68 ± 1.29
Socialist	46.54 ± 0.32
Radical Left	48.22 ± 0.77
Christian-Democratic	50.39 ± 0.45
Liberal	51.80 ± 0.52
Conservatives	52.01 ± 0.42
Radical Right	52.95 ± 0.65
Regionalist	52.96 ± 0.84
Green	58.63 ± 0.64
Confessional	58.77 ± 2.06

Table 2.10: Mean percentile bias score of families.

percentile bias. We also see that the parties with similar ideologies generally tend to have a similar bias - *Socialist* and *Radical Left* are close to each other, as are *Conservatives* and *Radical Right*.

We can also get the mean of the bias scores of the speeches of a parliamentarian to get the mean bias score for the parliamentarian. If we rank the parliamentarians by this score, we find certain extremist parliamentarians near the top and the president of the parliament close to the bottom. This is expected since the president needs to be non-partisan and generally uses objective language while speaking.

2.5.7 Bias in Social Media

We use the Climate Change Tweets dataset to test bias scoring in tweets. As tweets are short and many individual tweets do not contain any information, we combine each of *Anti*, *Pro* and *News* tweets into three documents and compute the bias score for these documents. *Anti* has the highest score of 31.63 followed by *Pro* with 31.17. This is expected since they typically express strong sentiments. *News* which is mostly factual reporting has the least score of 27.75, comparable to the mean score found for news articles in Section 2.5.4.

2.6 Summary

In this chapter, we developed a simple, interpretable model of bias in documents. The model combines elements of discrete-choice theory and word embeddings.

We curated two novel datasets based on Wikipedia revision histories to train and evaluate our model. The model is trained on pairs of revisions of Wikipedia articles, where one revision was corrected for POV issues by another edit. Formulating the problem

as identifying the larger bias in a pair, rather than estimating absolute bias, reduces subjectivity and issues of thresholding. We obtain strong performance on a holdout set of pairs of Wikipedia revisions.

Importantly, the model is interpretable: we can score individual words, a feature that an editor might rely upon to quickly identify the most problematic parts of a document that contribute to the bias. The list of globally most biased words contains a convincing list of strong adjectives and terms that tend to express emotions.

We explored the predictions of the model over datasets including news articles and law amendments. The bias distributions over the three domains (Wikipedia, news, laws) are quite different, with news the most biased, and laws the least, which can be explained by the policies governing the creation of content in each of them. We also observe that we can score the bias in different news outlets; these scores align well with crowdsourced labelings of bias of these outlets.

The model we developed can be integrated into applications to identify, measure, and monitor bias. For instance, one could build a browser extension to enable users to identify bias in online documents and thereby guard themselves against undue influence. Authors of documents that are expected to use objective language (such as legal documents or scientific articles) can measure the bias score to guide their writing[4]. Wikipedia and news editors could monitor bias as revisions are made to articles so as to take corrective action when needed.

Ultimately, we expect this work to contribute to better identifying and correcting both deliberate and subconscious bias in online discourse.

Broader Impact and Ethical Considerations. In addition to the bias scoring model we developed, which is applicable in a wide variety of domains, the methodology that we adopted of casting bias as a relative quantity and learning from pairwise comparisons can be extended to a much broader set of problems in natural language processing. It is particularly suited to those settings where the threshold for absolute categorization may be subjective or depends on many factors, while there is more agreement in comparisons. Examples include measuring hateful content, agreeableness, humor, sentiment, etc.

All data we use in this work is from publicly available sources. Wikipedia data that we collect is publicly released under the CC BY-SA and GFDL licenses and analysis of this content does not require informed consent.

Machine learning models are limited by the data that they learn from. Therefore our

[4]As an example, the bias score computed by our model for the abstract of this thesis is 14.84, which is relatively low (in between the mean scores for laws and Wikipedia articles). For comparison, the score for the acknowledgements, which contain subjective language, is 35.66, which is quite high (higher than the most biased news articles).

models inherit any bias that is inherent in Wikipedia's neutrality policy or the manner in which the editors interpret and enforce that policy. An editorial decision that is made based on the output of these models could also serve to reinforce such bias. However, the interpretability of our models mitigates this risk to some extent. For instance, if the model generates an unexpected output an editor can obtain the words that contributed to the model's assignment of a high or low bias score and perform an informed reassessment.

3 Social Media Campaigns

In this chapter,[1] we study the phenomenon of user engagement in social media campaigns, taking the case of communication about climate change on Twitter as an example.

With the goal of understanding effective strategies for communicating about climate change, we build interpretable models to rank tweets related to climate change, with respect to the engagement they generate. To rank the tweets, our models use a combination of the tweets' topic and metadata features. To remove confounding factors related to author popularity and minimize noise, the models are trained on pairs of tweets that are from the same author, are made around the same time period, and have a sufficiently large difference in engagement.

The models achieve good accuracy on a held-out set of pairs. We show that we can interpret the parameters of the trained model to identify the topic and metadata features that contribute to high engagement. We see that, among other observations, topics related to climate projections, human cost, and deaths tend to have low engagement, whereas those related to mitigation and adaptation strategies have high engagement. We expect the insights gained from this study will help craft effective communication about the climate in order to promote engagement, thereby lending strength to efforts to tackle climate change.

3.1 Introduction

Climate change is arguably one of the most important challenges facing humanity today with impacts on a global scale. Although surveys indicate that a majority of the population in many countries is now knowledgeable about climate change and its effects, the level of awareness and support for climate-friendly policies still vary widely among

[1] This chapter is based on (Suresh, Milikic, et al., 2023). The author of this thesis led the project, performed dataset curation, designed the model and experiments for evaluation, and interpreted the model and results.

people from different countries, income and education levels, and age groups (Flynn et al., 2021; Leiserowitz et al., 2022). It is important to use communication to spread awareness and to promote engagement among the groups that are less involved; in order to further increase the scale of actions towards mitigation and adaptation, and to pressure lawmakers and governments to create and enforce climate-friendly laws.

There are several subtopics within the climate change issue, including the causes, effects, and mitigation and adaptation strategies. The potential to create engagement is likely to be different for each of these subtopics. For instance, complex technical details might be less appealing than vivid descriptions of the effects or of promising solutions. Our goal in this work is to discover in a data-driven way the strategies of communication, as defined by the subtopics being emphasized, that are more effective in creating engagement among a general audience.

The data source that we use for this study is Twitter. Twitter is extensively used for communication about climate change by individuals, activist groups, and government agencies. Each tweet is associated with several metrics, such as the number of *likes*, *retweets*, and *replies*, that quantify the engagement that it generates. All tweets, along with their metadata, are publicly available and can be easily accessed through the Twitter API[2]. This enables us to curate a large and rich dataset on which to train interpretable engagement prediction models.

However, a significant challenge in building such models is the presence of confounding factors such as author popularity. A tweet might generate strong engagement because its author is popular rather than because of its engaging content. Another potential confounder is the change over time in public interest in climate change. For instance, tweets about climate change made around the time of extreme weather or a major climate change conference might receive greater attention than tweets at other times. We minimize the effect of such confounders by defining the task as comparing the engagement within a pair of tweets rather than predicting the engagement for a given tweet. The pair of tweets are chosen to be from the same author and from the same window in time.

The chapter is organized as follows. In Section 3.2, we give a short description of relevant prior work. In section 3.3, we describe the dataset of tweets that we curated. The features we use for prediction and our model architecture are given in Section 3.4. We describe training and evaluation and discuss our results in Section 3.5, and we conclude the chapter in Section 3.6.

[2]The Twitter API changed recently and no longer provides this level of access for free.

3.2 Related Work

Communication about climate change is an active area of research and the question of which strategies are best to promote engagement has been well-studied in the area (Agin & Karlsson, 2021). Gustafson et al. (2020) find that sharing personal stories on the radio about the harmful effects of climate change can be a persuasive strategy. Xia et al. (2021) analyze the spreading behavior of climate-related tweets and identify the factors responsible for virality, from a dataset obtained through manual coding of a small number of tweets.

Tweet engagement prediction in general (not restricted to climate change) is also a well-studied problem. Tan et al. (2014) study the effect of wording on engagement by comparing pairs of tweets from the same user about the same topic. More recently, K. Wang et al. (2018) used multimodal information (images and text) to predict retweet behavior. Topic modeling (inferring latent topics in text) has been used in Dahal et al. (2019) to understand which topics are prevalent in the global climate-change discourse. To the best of our knowledge, we are the first to apply topic modeling in a pairwise comparison framework to predict engagement in tweets about climate change and to interpret the learned model in order to understand the engaging subtopics.

3.3 Dataset

We use the Twitter API[3] (Twitter, 2023) to obtain 8,041,921 tweets that are related to climate change and were created between January 1st, 2021 and November 4th, 2022. To decide whether a tweet is related to climate change, we check if it contains one of the keywords in the 'General' topic category of a taxonomy for studying climate-change tweets (UN Global Pulse, 2014),. We keep only the tweets in English (94.28% of the dataset). For each tweet, we keep its full text, author, and information about whether it contains URLs, hashtags, animated GIFS, images, or videos. We also keep the public engagement metrics about the tweet, i.e., the number of likes, retweets, and replies that it obtained.

We then construct the pairs of tweets to be compared. For each author, we review their history and obtain the pairs of tweets that were created within seven days of each other and that have a difference in engagement (sum of likes, retweets, and replies) of 100 units or 10%, whichever is higher. This method of creating pairs ensures that we avoid confounding factors related to the author and time, and that we minimize the noise in the comparison. We finally end up with 774,507 pairs of tweets that we use for training our engagement prediction model.

[3]The Twitter API changed recently and no longer provides this level of access for free.

3.4 Model

Our model uses both the topic of the tweets in a pair, as well as their metadata features, to predict which of the two tweets receives greater engagement.

3.4.1 Features

To infer the topics present in a tweet, we use the TopClus method introduced by Meng et al. (2022). The method uses the pre-trained language model BERT (Devlin et al., 2019) in combination with an attention model for representing texts (tweets in our case). It then uses a latent variable model to cluster these representations into topics in an unsupervised fashion. The learned topics can be interpreted by looking at the tweets whose representations are closest to the center of the clusters. Once the model is trained on a set of tweets, it can be used to infer the probability that a new tweet belongs to each of the topics (i.e., to compute the tweet's topic probability distribution). The number of clusters or topics K is a parameter that needs to be set prior to training the TopClus model. More details about TopClus are given in Section 1.3.

In order to have a good representation of the tweets for clustering, we pre-process their texts to be similar to the texts used for training the BERT model by dropping URLs, mentions, and hashtags. Due to computational constraints, we train the TopClus model on a 10% random sample of our dataset, instead of on the full dataset. We then use the trained model to infer the topic probability distributions of all the remaining tweets in our dataset. Thus, in the end, we have for each tweet i a topic feature vector[4] $\mathbf{t_i} \in \mathbb{R}^K$.

In addition to topic features, we also include metadata features for each tweet. These are represented as binary values (1 if the corresponding feature is present, 0 if not), and capture the presence of URLs, hashtags, animated GIFs, images, and videos. We also include a binary value to represent if the tweet was created during working hours (weekdays, 9 am-5 pm) on the East Coast of the US which is likely to have a large fraction of the audience, namely English-speaking Twitter users. We represent the metadata features of tweet i by $\mathbf{m_i} \in \mathbb{R}^6$.

3.4.2 Model Architecture

The model that we use is reminiscent of the Bradley-Terry model of pairwise comparison outcomes (Bradley & Terry, 1952). We define the probability that tweet i in a pair achieves more engagement than the other tweet j to be

[4]We use $K = 100$ as it gave interpretable topics. It is also the default value suggested in Meng et al. (2022).

$$P(i \succ j) = \frac{e^{s_i}}{e^{s_i} + e^{s_j}}, \tag{3.1}$$

where $s_i, s_j \in \mathbb{R}$ are the *engagement potentials* of tweets i and j respectively. The higher the engagement potential of a tweet, the more likely it is to have higher engagement in a pair.

We model the engagement potential as a linear function of the topic and metadata features. More precisely, we define

$$s_i = \mathbf{w_t}^T \mathbf{t_i} + \mathbf{w_m}^T \mathbf{m_i}, \tag{3.2}$$

where $\mathbf{w_t} \in \mathbb{R}^K$ and $\mathbf{w_m} \in \mathbb{R}^6$ are the learned vectors of coefficients for topics and metadata respectively. The choice of a linear model enables us to easily understand, by interpreting the coefficient values, the effect of the features hence to find effective strategies that are correlated with high engagement[5]. We train three versions of our model; they differ in the features they include: *Meta* (only $\mathbf{m_i}$), *Topic* (only $\mathbf{t_i}$) and *Topic+Meta* (both $\mathbf{t_i}$ and $\mathbf{m_i}$).

However, note that the model, as defined so far, is not identifiable with respect to the topic coefficients $\mathbf{w_t}$ (we cannot precisely estimate their values). In fact, if we modify the model by adding a constant to each of these coefficients, we obtain the same predictions. This can be seen by noting that the sum of the components of topic vectors $\mathbf{t_i}$ are always one (as they are probability distributions). To avoid this issue we fix one of the coefficients to be zero.[6]

3.5 Experiments and Results

We split our dataset of tweet pairs chronologically[7] into training, validation, and testing sets, at the ratio 90:5:5. We train our models by finding the optimal set of coefficients $\mathbf{w_t}$, $\mathbf{w_m}$ that maximize the probability of the pairs in the training set under the model in Equation 3.1. We use classification accuracy as the evaluation metric, because our dataset is balanced and errors are symmetric.

[5]We also tried slightly more sophisticated models like the quadratic form in Chapter 2. However, these didn't improve performance, presumably because the topic features are sparse, unlike the embedding vectors.

[6]The choice of this coefficient does not affect the relative ordering of coefficients and hence their interpretation.

[7]We split chronologically rather than randomly as that is closer to a real setting where we would want to predict the engagement via future tweets based on past tweets.

We do not use regularisation, as overfitting is unlikely (our model is simple and we have a large amount of data). Indeed, we see from the validation performance that overfitting does not occur. We re-train on the union of the training and validation sets, and we report our accuracy on the test set (see Table 3.1). For comparison, we also report the human accuracy obtained by one of the authors who labeled 200 random pairs from the test set.

Table 3.1: Test accuracies of different models with 95% confidence intervals

Random	**Meta**	**Topic**	**Topic+Meta**	**Human**
50.17 ± 0.50	58.90 ± 0.49	64.54 ± 0.48	$\mathbf{66.53 \pm 0.47}$	65.00 ± 6.61

We see that all our models significantly outperform the random baseline and achieve performance comparable to a human. *Topic* outperforms *Meta* and the best performance of 66.53% is achieved by *Topic+Meta*, suggesting that the two features provide complementary information. Tan et al. (2014) report a similar accuracy for models based on wording and metadata, when predicting relative engagement via tweets containing the same URL. We also tried models that use the words as features, instead of topics. The accuracy was slightly lower, and the interpretation was more difficult for those models, as a clear pattern could not be seen among the most predictive words. The accuracy for word-based models could possibly be increased by using contextual word embeddings and state-of-the-art sequence models such as Transformers (Vaswani et al., 2017), but their interpretation is likely to still be difficult.

We re-train *Topic+Meta* on the complete set of pairs to obtain the coefficients for interpretation. We show coefficients of the top 24 topics in terms of their prevalence in our data in Figure 3.1; these are also the most interpretable. A higher coefficient indicates more engagement. The names of the topics were manually assigned by examining the top 500 tweets with the highest probability for the topic. A random sample of 10 tweets from this set for each of the topics in the table is given in Appendix A for a more detailed interpretation. The coefficients of the metadata features are given in Table 3.2, where a positive coefficient is correlated with higher engagement.

Table 3.2: Coefficients of metadata features with 96% confidence intervals

URL	**Hashtag**	**GIF**	**Video**	**Image**	**WorkHr**
-1.54 ± 0.13	-0.12 ± 0.12	0.47 ± 0.28	0.76 ± 0.13	0.58 ± 0.10	-0.24 ± 0.10

Topics without much useful content, such as *Links/Promo* clearly have low engagement. Tweets about the long-term *Projections* of climate-change effects also have low engagement, which could be due to the temporal discounting of risks that are in the distant future. Topics such as the *Human cost* of climate change and *Deaths* have low engagement, as do the topic of stocks and *Investment*.

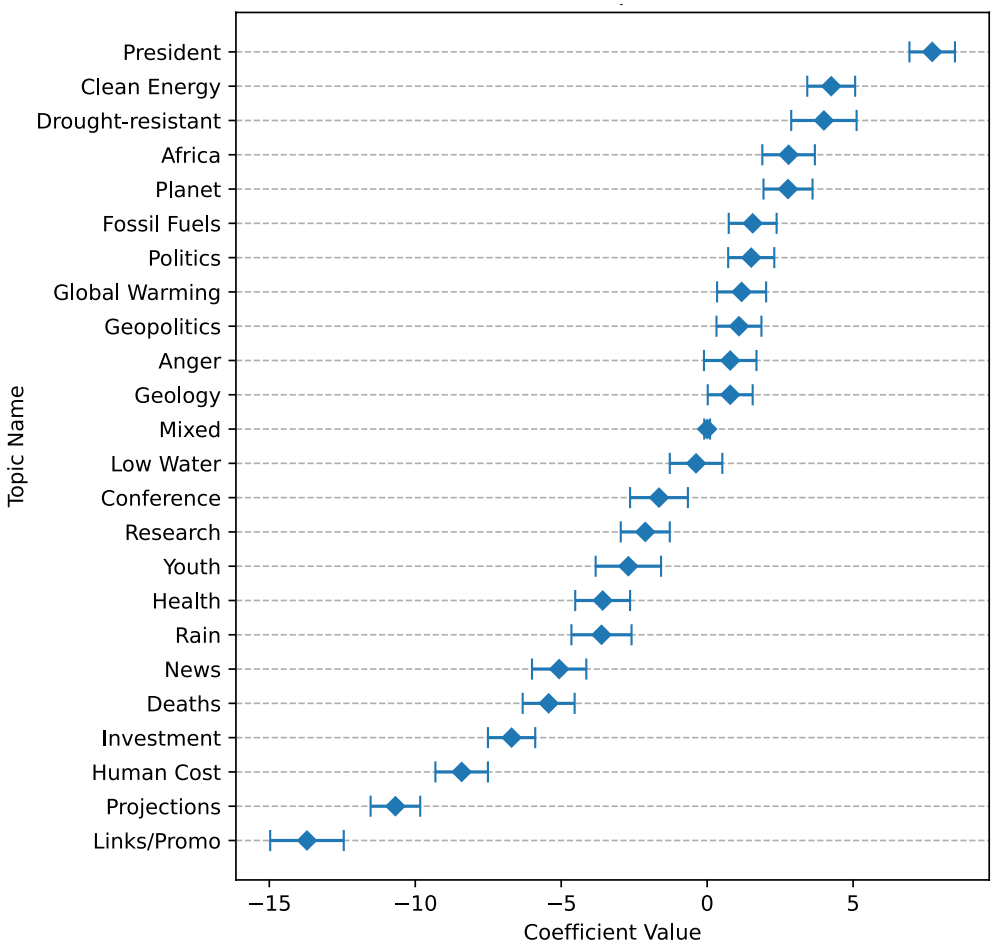

Figure 3.1: Topic coefficients with 96% confidence intervals

Topics that contribute to high engagement include discussions about the *President*'s actions, and potential strategies for adaptation and mitigation such as *Drought-resistant* plants and *Clean Energy*. Interestingly, tweets about the effects of climate change in *Africa* receive high engagement, as do those about climate change on other *Planets* and in Earth's geological history. Tweets about climate change *Conferences* and *Research* have a moderate level of engagement.

Looking at the coefficients for metadata features, we can see that tweets that include URLs tend to receive less engagement (probably due to less information in the tweet itself). As could be expected, tweets created during working hours also receive less engagement. Whereas, tweets that contain animated GIFs, videos and/or images tend to have higher engagement.

3.6 Summary

In this chapter, we curated a large dataset of tweets related to climate change and built interpretable predictive models of tweet engagement. We avoid confounding factors by formulating the task as a pairwise comparison of engagement among a pair of tweets that are from the same author and are emitted around the same time. The models achieve good accuracy on a held-out set of pairs. By interpreting the coefficients of the models, we could discover the topic and metadata features that are correlated with high engagement. This information could conceivably be used to guide the creation of new climate-related tweets and other communication about the climate in order to promote engagement among the population, thereby giving strength to citizen-driven efforts for tackling the issue of climate change. Finally, it is worth noting that while we focused on communication about climate change as an example, the methods we developed in this chapter can be applied to identify effective engagement strategies for social media campaigns in general.

4 Lobbying

In this chapter,[1] we present an NLP-based method for studying the influence of interest groups (lobbies) in the law-making process in parliaments, taking the European Parliament (EP) as an example.

We collect and analyze novel datasets of lobbies' position papers and speeches made by members of the EP (MEPs). By comparing these texts on the basis of semantic similarity and entailment, we are able to discover interpretable links between MEPs and lobbies.

In the absence of a ground-truth dataset of such links, we perform an indirect validation by comparing the discovered links with a dataset, which we curate, of retweet links between MEPs and lobbies, and with the publicly disclosed meetings of MEPs. Our best method achieves an AUC score of 0.77 and performs significantly better than several baselines. Moreover, an aggregate analysis of the discovered links, between groups of related lobbies and political groups of MEPs, correspond to the expectations from the ideology of the groups (e.g., center-left groups are associated with social causes).

We believe that this work, which encompasses the methodology, datasets, and results, is a step towards enhancing the transparency of the intricate decision-making processes within democratic institutions.

4.1 Introduction

The transparency of decision making is of central importance for the legitimacy of democratic institutions such as parliaments. The influence of interest groups (lobbies) on parliamentarians and the potential for a resultant subversion of the power of the electorate to determine policy have led to demands from groups, such as Transparency International

[1]This chapter is based on Suresh, Radojevic, et al. (2023). The author of this thesis led the project, performed dataset curation, designed the models and experiments for evaluation, and interpreted the model and results.

(1993), for effective rules and systems to increase transparency. The emergence of several open government initiatives around the world (European Union, 2021; Obama White House, 2018; Swiss Government, 2021) is in part a response to such demands.

The EU Transparency Register (TR) (European Union, 2011) is one such initiative that provides a tool for EU citizens to explore the influence of interest groups in the European Parliament (EP). Any organization that seeks to influence EU policy, with a few notable exceptions, needs to register with the TR before meeting with parliamentarians. The organizations are asked to disclose information such as their address, website, financial information, and goals.

However, the EU TR has several limitations. The disclosure of most of the information is voluntary and there is little oversight. It is difficult to obtain information regarding which members of the EP (MEPs) or laws are targeted (and by which particular lobbies) and to know the lobbies' positions on specific policies.

There have been several studies conducted by the political science community on EU lobbying (Bouwen, 2003; Rasmussen, 2015; Tarrant & Cowen, 2022). However, these studies focus either on a single policy issue or on a small set of issues, and/or they are limited in terms of sample size as they employ less scalable methodologies such as manual examination of position papers and individual interviews.

One exception is a study by Ibenskas and Bunea (2021). They analyze the Twitter follower network of a large number of MEPs and lobbies from the TR, with respect to the MEP's nationality and committee memberships and lobbies' self-reported interests in the TR. However, they do not analyze the textual content of MEPs' speeches and amendments and the lobbies' position papers, which would be instrumental for uncovering convergence on specific policy issues beyond the broad interest areas mentioned in the TR.

Therefore, there is a need for automated approaches for studying lobbying in a comprehensive manner, with the help of rich publically available textual resources and by using modern tools developed by the NLP community. In particular, self-supervised algorithms for text representation and computing text similarity and entailment (Reimers & Gurevych, 2019a) are promising for identifying interesting patterns. A major challenge faced by such automated approaches, even unsupervised ones, is the lack of ground-truth data for validation. As far as we are aware, there exists no large database of verified MEP-Lobby links, let alone one annotated for relevant policy positions.

In this chapter, we present an NLP-based approach for automatically discovering potential links between a large number of MEPs and lobbies, by comparing the text in publicly available documents where they express their views on policy issues. We define a link between an MEP and a lobby as a convergence of views between them on some policy issue. To the best of our knowledge, such an approach has not been explored in prior

work. We focus on the eighth term of the EP (2014-2019), as it was the last complete term that was not disrupted due to the pandemic. In the absence of ground-truth data, we perform an indirect validation by comparing the discovered links to a dataset we curate of retweet links between MEPs and lobbies.

We use the retweet network instead of the follower network studied by Ibenskas and Bunea (2021), because retweets typically occur as a result of the agreement of particular views between the MEP and lobby, in contrast to 'follows' that can result from a general interest in knowing more about a topic or person (Metaxas et al., 2015). Moreover, timestamps for retweets are publicly available, which allows us to collect more relevant data for the eighth term.

Our methods are also designed to be interpretable - we can obtain the specific set of MEP speeches and lobby documents that match for an MEP-Lobby pair, thus enabling manual validation of discovered links by users.

Since 2019, it has been mandatory for MEPs in certain key positions (such as reporters of parliamentary committees) to publish their meetings with lobby groups (European Parliament, 2019). We use this data as an additional source of validation, although it only covers the subset of the MEPs from the eighth term who were re-elected in the ninth term.

The chapter is structured as follows. In Section 4.2, we describe the datasets that we curate and use. In Section 4.3, we describe the different methods that we experiment with for discovering links. We evaluate the methods in Section 4.4 and interpret them in Section 4.5. We conclude the chapter in Section 4.6.

4.2 Datasets

We curate several novel datasets for our study. To obtain the policy positions of lobbies, we curate a dataset of position papers (Section 4.2.1). The views of the MEPs are obtained through a dataset of their plenary speeches (Section 4.2.2) and proposed amendments (Section 4.2.2). For validation, we use a dataset of MEP-lobby retweet links (Section 4.2.3) and meetings (Section 4.2.3).

4.2.1 Lobbies

Our data collection pipeline for lobbies is given in Figure 4.1. The versions of the data at different stages of the pipeline are labeled as $D1$, $D2$, and so on. Information on the size of these datasets is given in Table 4.1. We now describe the steps in the pipeline.

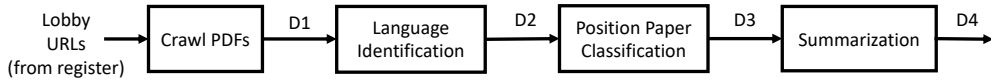

Figure 4.1: Data collection pipeline for lobbies. $D1$ contains all crawled PDF documents, $D2$ contains all English documents in $D1$, $D3$ contains the documents in $D2$ classified as position papers, and $D4$ contains the summaries of the documents in $D3$.

Table 4.1: Lobby datasets

	$D1$	$D2$	$D3, D4$
Documents	766,437	373,216	48,970
Lobbies	4,230	3,965	2,558

Crawling and Language Identification

We focus on the lobbies that were on the EU TR under the heads of *Trade and Business Associations*, *Trade Unions and Professional Associaions* and *Non-Governmental Organisations*, as of October 2020; this is a total of 5,461 lobbies. Although some other categories like *Companies and Groups* are also influential, we do not include them because they are mostly represented by associations that they are part of and rarely publish position papers of their own.

We obtain the URLs of the lobby websites from the TR and crawl publicly available PDF documents from them to obtain an initial dataset $D1$. We parallelize the crawling by using HTCondor (HTCondor, 2023) on a cluster of 300 nodes with maximum limits of 250 MB of text and 5 hours of crawling per website and are able to crawl all PDFs in nearly 70% of the lobby websites in about four days. To identify the languages in the dataset, we use the Fasttext language identification model (Joulin, Grave, Bojanowski, Douze, et al., 2016; Joulin, Grave, Bojanowski, & Mikolov, 2016). As nearly half (48.7%) of all the documents are in English and other languages appear in much smaller percentages, we keep only the English documents ($D2$) to simplify the rest of the analysis.

Position Paper Classification

A large majority of the PDFs do not contain significant information about lobby policy positions, including documents such as product brochures, user manuals, technical documentation, forms, etc. By manually labeling 200 randomly sampled PDFs, we estimate the proportion of PDFs that contain policy positions to be approximately 22.5%. In order to reduce noise in the data and to enable us to apply methods that are more performant but less scalable, we classify the PDFs into *position papers* and *other documents* and work with those classified as position papers.

We train a weakly supervised logistic regression model by using TF-IDF features for this task and by using the presence of the word 'position' in the URL as the label. On a manually labeled validation set of 200 PDFs, the model achieves a precision of 95% and a recall of 39% in identifying position papers. The most predictive words include *position*, *should*, *strongly*, etc. and are indeed likely to be present in texts articulating positions. We then apply the classifier on all PDFs in $D2$, and keep those that are classified as position papers to obtain $D3$.

Summarization

Many of the documents are quite long (greater than 1,000 words) and cannot be encoded fully by pre-trained encoders such as SentenceBERT (Reimers & Gurevych, 2019a). They also typically contain information, such as technical details, that is not relevant for matching with MEP speeches. Hence, we summarize the documents into three-to-four sentences that capture the main ideas expressed. This also makes the interpretation of matched document-speech pairs easier.

We experiment with various state-of-the-art summarization models and find through manual examination that OpenAI's `gpt-3.5-turbo` (the model behind ChatGPT) (OpenAI, 2023a) generates coherent summaries that capture the most salient points expressed in the document. We thus generate the dataset $D4$ that contains the GPT-generated summaries of documents in $D3$. We summarize only the documents in $D3$, despite the low recall of position paper classification due to the cost constraints of using the ChatGPT API.

Lobby Clustering

Individual lobbies are so numerous and specialized that it is difficult to see interpretable patterns, even after a successful MEP-Lobby matching. We, therefore, cluster the lobbies into relatively homogenous groups by using the description of their goals in the EU TR. We first convert these descriptions to short phrases (3-4 words) by using ChatGPT and cluster the phrases by using K-Means[2] after embedding them using SentenceBERT. The clusters are mostly straightforward to interpret, although some of them contain a few unrelated lobbies. Some clusters, related particularly to energy, include both renewable energy companies and fossil-fuel companies. This is probably because some of the fossil-fuel companies are undergoing a renewables transition and emphasize this in their goal statements in the TR.

Finally, we ask ChatGPT to name each of the clusters, based on the short phrase descriptions of the lobbies in each of them. Most of the names are highly representative and specific, but a few of them are too generic (e.g., 'Interest groups in the EU'); we

[2]We use $K = 100$ as it gives mostly coherent clusters with minimal duplicates.

then manually correct the overly generic ones to make them more specific. The list of the top three lobby clusters with the most position papers is given in Table 4.2 with a couple of examples of lobbies that are in each cluster. The full list of lobby clusters with their lobbies is given in Appendix B. The numbers after the cluster names in Table 4.2 serve to differentiate clusters with identical names and correspond to the serial number in the complete table in Appendix B.

Table 4.2: Top three lobby clusters by number of position papers. All three have about 1,400 papers each.

Lobby Cluster	**Example Lobbies**
Manufacturing - 50	orgalim.eu glassforeurope.com
Renewable Energy - 45	solarpowereurope.org windeurope.org
Business - 8	enterprisealliance.eu smeeurope.eu

4.2.2 MEPs

Data on MEPs' policy positions are obtained from two sources: their speeches in the plenary sessions of the EP, and the law amendments that they propose within parliamentary committees. We describe each of them in the following sections.

Speeches

We scrape all plenary speeches of the eighth term from the EP website (51,432 in total), spoken by 849 MEPs (and a few non-members). The speeches are organized into 1,471 debates with titles; each debate is about a specific law or policy issue. For the speeches made by MEPs, we scrape the official EP ID of the MEP, which we use to query the Parltrack database (Parltrack, 2023) to obtain additional information about the MEP, such as their name, nationality, party, etc.

Similar to the case for lobbies, it is easier to find patterns if we analyze the links to lobbies for *groups* of MEPs rather than individuals. MEPs are naturally grouped according to their ideology into nine political groups. The European People's Party (EPP, center-right) and the Socialists and Democrats (S&D, center-left) are the two largest groups.

To quantify the ideological position of the groups, we use data from the Chapel Hill Expert Survey (CHES) (Jolly et al., 2022), where political scientists have scored every party on a numerical ideological scale ranging from zero (extreme left) to ten (extreme right). In addition to the general left-right ideology (which are referred to simply as

'Ideology'), the survey also contains scores for more fine-grained aspects of ideology such as views on how to manage the *economy* (state control vs. free market), views on *social* issues (libertarian vs. traditional/authoritarian), and views on *EU* integration (anti-EU vs. pro-EU)[3]. We aggregate the party-level data from CHES to get the scores for the political groups[4]. The positions of the nine EP groups are given in Table 4.3.

Table 4.3: Political groups and ideology scores, sorted by general left-right ideology.

Group name	**Acronym**	**Ideo**	**Econ**	**Soc**	**EU**
Confederal Group of the European United Left - Nordic Green Left	GUE/NGL	1.65	1.39	3.31	3.49
Group of the Greens/European Free Alliance	Greens/EFA	3.21	3.22	2.21	5.61
Group of the Progressive Alliance of Socialists and Democrats in the European Parliament	S&D	3.83	3.90	3.83	6.18
Group of the Alliance of Liberals and Democrats for Europe	ALDE	6.09	6.70	4.00	6.05
Europe of Freedom and Direct Democracy Group	EFDD	6.55	5.43	5.63	1.40
Group of the European People's Party (Christian Democrats)	EPP	6.69	6.32	6.38	5.89
European Conservatives and Reformists Group	ECR	7.21	5.90	7.28	3.33
Europe of Nations and Freedom Group	ENF	9.32	6.14	8.89	1.31
Non-attached Members	NI	9.76	4.06	9.54	1.18

The speeches for the eighth term are available only in the original language of the speaker, unlike in earlier terms where the EP provided translated versions in all official EU languages, including English. Hence, we automatically translate all the non-English speeches to English by using the open-source OPUS-MT models (Tiedemann & Thottingal, 2020) provided in the EasyNMT package (Reimers, 2022). Apart from the dataset of full speeches, we also generate a dataset of the speech summaries by using ChatGPT as in Section 4.2.1.

[3]The CHES codebook refers to these scores as LRGEN, LRECON, GALTAN, and EU_POSITION

[4]We take the weighted average of party scores with weights being the size of each party in the group.

Amendments

We use the law amendments dataset released by Kristof et al. (2021), which contains 104,996 amendments proposed by MEPs in the eighth term on 347 laws identified by their titles. We input the old and new versions of the law articles changed by the amendment to ChatGPT, along with the law title, and ask it to generate a possible sentence for the position paper of a lobby that would like to get this amendment accepted. We expect to be able to match the sentence to the lobby summaries generated in Section 4.2.1.

We find that ChatGPT generates a concise summary of the amendment's effect on the law, adding that it (by impersonating a lobbyist) would like such a change to be effected. Short but significant changes to the law are correctly interpreted by the model, such as the change from *shall* to *should* being a change from a mandatory requirement to a recommendation. However, the model tends to "hallucinate" when there is insufficient context, such as in the case of entire articles being deleted or new ones being added. Therefore, we restrict this procedure to generate summaries exclusively for the 88,853 amendments that only modify existing articles without deleting them entirely.

4.2.3 Validation Datasets

For validating the discovered links, we curate a dataset of retweet links and a dataset of MEP-Lobby meetings. We describe them in the following sections.

MEP-Lobby Retweet Links

We obtain the Twitter handles of MEPs from multiple sources including official profile pages on the EP website, the Parltrack database, other third-party databases, and manual search. We were able to obtain handles for 669 MEPs. We collect handles of the lobbies with position papers by scraping their homepages for 'Follow us on Twitter' links, and obtain 1,676 handles. We see that, indeed, most of the MEPs and Lobbies have a presence on Twitter.

Once we have the handles, we use the Full Archive Search endpoint of the Twitter API[5] (Twitter, 2023) to retrieve the content and metadata of all their public tweets during the period of the eighth term. We then identify the tweets of an MEP (resp. lobby) that are 'pure' retweets (without any added original content hence less likely to indicate disagreement) and check if the referenced tweet is from a lobby (resp. MEP). We consider that there is an (undirected) retweet link between an MEP-Lobby pair if either the MEP or the lobby has retweeted the other at least once, which leaves us with 8,754 links.

[5]The Twitter API changed recently and no longer provides this level of access for free.

MEP-Lobby Meeting Links

Data on meetings between MEPs and lobbies since the beginning of the ninth EP term (2019-2024) are available from the Integrity Watch Data Hub (Integrity Watch, 2023). Integrity Watch monitors and collects meeting information from the European Parliament website. Every meeting includes an MEP identified by the EP ID and a list of lobby names or acronyms. We match the lobby names to our data from the register using fuzzy string matching, thus enabling us to establish 1,365 links between 125 MEPs from the eighth term (who were re-elected in the 9th term) and 565 lobbies.

4.3 Methods

Here, we describe the framework and methods we use to discover links between MEPs and lobbies. Let $\mathcal{M}$ denote a set of MEPs and $\mathcal{L}$ denote a set of lobbies. We assume that an MEP $m \in \mathcal{M}$ and a lobby $l \in \mathcal{L}$ have a link with some probability $P(m, l)$. One possible approach to discovering links is to estimate this probability directly. However, this is difficult as we do not have a ground-truth dataset of links on which to train a probabilistic model. Without such data, we can make only *relative* assessments of $P(m, l)$, based on information about the similarity of views between m and l. Thus, we can say that $P(m_1, l_1) > P(m_2, l_2)$ if the similarity of views for the pair (m_1, l_1) is higher than that for the pair (m_2, l_2).

Hence, we adopt the following framework. Given an MEP $m \in \mathcal{M}$, and a lobby $l \in \mathcal{L}$, the goal of our methods is to compute an association score $A(m, l) \in \mathbb{R}$ such that

$$A(m_1, l_1) > A(m_2, l_2) \iff P(m_1, l_1) > P(m_2, l_2) \ \ \forall m_1, m_2 \in \mathcal{M}, \ \ \forall l_1, l_2 \in \mathcal{L}. \quad (4.1)$$

The methods differ in how $A(m, l)$ is computed. For methods using texts, we use $\mathcal{S}_m$ to refer to the documents produced by m and $\mathcal{D}_l$ for the documents produced by l.

4.3.1 Baselines

We first describe the baselines. The goal of comparing our models to these baselines is to check if the content of the texts provides non-trivial information about the MEP-Lobby association.

Random

This is the simplest baseline where we have $A(m, l) \sim \text{Uniform}(0, 1)$.

Prolificacy (Pr)

This baseline is based on the intuition that the MEPs and lobbies that are more prolific and generate more texts are more likely to have a link between them. Hence, for this baseline, we define $A(m, l) = |\mathcal{S}_m| \times |\mathcal{D}_l|$.

Nationality (Nat)

Prior work suggests that there is a strong tendency for MEPs to associate with lobbies from the same EU member state (Ibenskas & Bunea, 2021). We therefore include a baseline where $A(m, l) = 1$ if m and l are from the same member state and $A(m, l) = 0$ otherwise.

4.3.2 Text-Based Methods

Here, we describe our methods that use the content of the texts in $\mathcal{S}_m$ and $\mathcal{D}_l$.

Text Classification (Class)

We train a fastText (Supervised) classifier to predict whether a given text was generated by a particular lobby. We use the sentences in the lobby dataset $D2$ for training the classifier. Independent linear classifiers are trained for each lobby, but they share the same embedding layer, which enables the model to scale to a large number of classes while having limited data for each class. The linear structure allows interpretability; the top predictive words for some lobbies are given in Table 4.4. We see that these clearly reflect the areas of work of the lobbies.

amnesty.eu	executions	detainee	occupants	assurances	reassignment
businesseurope.eu	globalisation	kyoto	relocation	lisbon	wto
caneurope.org	climate	warming	fossil	coal	allowances
fuelseurope.eu	refineries	refinery	gasoline	fuels	cis
ficpi.org	invention	trademarks	patent	practitioner	attorneys
orgalim.eu	manufacturers	machines	engineering	doc	counterfeiting

Table 4.4: Top predictive words for some prominent lobbies

Once the classifier is trained, we compute

$$A(m, l) = \frac{1}{|\mathcal{S}_m|} \sum_{s \in \mathcal{S}_m} P(l|s), \tag{4.2}$$

where $P(l|s)$ is the probability that lobby l generated the text s, according to the trained classifier.

Semantic Similarity (SS)

In this method, we first convert the texts in $\mathcal{S}_m$ and $\mathcal{D}_l$ to vector representations that capture their meaning. The cosine similarity between these vectors gives a measure of semantic similarity between the texts.

We use the pre-trained `all-MiniLM-L6-v2` model from SentenceBERT to obtain 384-dimensional vector representations for the texts. We then compute

$$A(m,l) = \max_{s \in \mathcal{S}_m, d \in \mathcal{D}_l} \mathbf{v_s}^T \mathbf{v_d}, \tag{4.3}$$

where $\mathbf{v_s}$ and $\mathbf{v_d}$ are the vector representations of texts s and d respectively, normalized to unit norm.

If the whole text fits within the maximum sequence length for the SentenceBERT model (256 tokens), it is encoded into a vector directly. This is the case for summary texts. If the text is too large to fit, we separate it into individual sentences and take the normalized sum of the sentence encodings.

Entailment (Ent)

One issue with SS is that there exist cases where two texts contradict each other, but they still have high semantic similarity based on their vector representations. This can cause false positives in the discovered links. For instance, an MEP's speech about increasing a specific tax could be matched with a lobby's position paper advocating for a reduction of the same tax. One reason for this is that the fixed-length vector representation might not always have enough information to process negations.

In order to reduce such cases, we use a cross encoder model pre-trained on natural language inference (NLI) data, including SNLI and MultiNLI. We use, in particular, the `cross-encoder/nli-deberta-v3-base` model from SentenceBERT. Given a pair of texts (s, d), this model is trained to output whether s contradicts d, s entails d, or neither.

As texts from an MEP speech and lobby document are usually less similar than a pair of premise and hypothesis from NLI, the model assigns the highest probability to *neither* for most of the text pairs. However, we can identify probable contradictions, especially for highly similar pairs, by checking if the probability it assigns to *contradiction* ($P(con)$) is greater than that for *entailment* ($P(ent)$).

We then compute

$$A(m,l) = \max_{s \in \mathcal{S}_m, d \in \mathcal{D}_l} \mathbf{v_s}^T \mathbf{v_d}, \quad \text{s.t} \;\; P_{(s,d)}(ent) > P_{(s,d)}(con). \tag{4.4}$$

4.4 Evaluation

We evaluate our methods on both the retweet links and meetings datasets. We use the area under the receiver operating characteristic (ROC) curve (AUC), as our metric because it is independent of the choice of a threshold for $A(m,l)$. We are mostly interested in the low false positive rate (FPR) regime of the ROC as we expect the MEP-Lobby influence network to be sparse. Hence, we also compute the partial AUC (pAUC) for the FPR < 0.05 region.

The scores of all methods are given in Table 4.5 and the ROC curves for retweets and meetings are in Figure 4.2 and Figure 4.3 respectively. We denote in parentheses the documents used for the sets $\mathcal{D}_l$ (D2:all English documents, D3:Position Papers, D4:Summaries) and $\mathcal{S}_m$ (Sp.:Speeches, Amd: Amendments). For a fair comparison between methods, the evaluations include only the set of lobbies that have position papers.

Method	Retweets		Meetings	
	AUC	pAUC	AUC	pAUC
Random	0.500	0.025	0.500	0.025
Pr(D2,Sp.)	0.603	0.052	0.598	0.048
Pr(D3,Sp.)	0.652	0.092	0.673	0.111
Pr(D2,Amd)	0.605	0.059	0.668	0.070
Pr(D3,Amd)	0.646	0.106	0.724	0.150
Nat	0.530	0.076	0.551	0.107
Class(Sp.)	0.687	0.079	0.652	0.070
SS(D2,Sp.)	**0.763**	0.189	0.751	0.147
SS(D3,Sp.)	0.742	0.185	0.759	0.156
SS(D4,Sp.)	0.759	0.196	**0.780**	0.176
SS(D4,Amd)	0.704	0.169	0.773	**0.208**
Ent(D4,Sp.)	0.758	**0.198**	0.774	0.175

Table 4.5: Evaluation results of baselines (top half) and our methods (bottom half). The pAUC is computed on the region where FPR ≤ 0.05.

We clearly see that the text models that use semantic similarity and entailment outperform all baselines and the text classification model on both datasets. In fact, the classification model is worse than some baselines. We think this could be because it is unable to capture all aspects of a lobby's position in the fixed-length classifier weights, while the similarity-based methods do not have this constraint. Summarization seems to help in

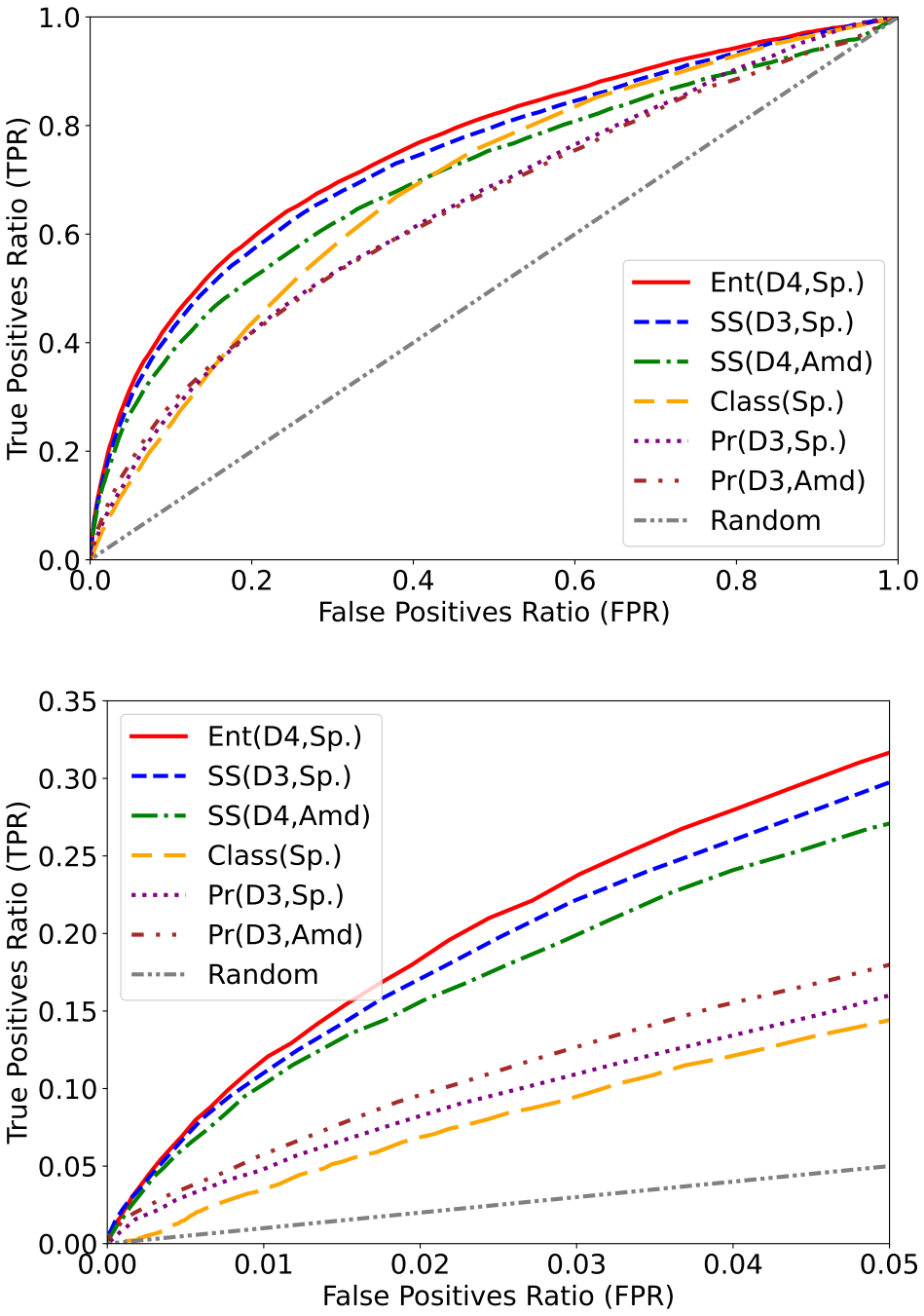

Figure 4.2: ROC curves for the Retweet dataset - Full (Top) and FPR≤0.05 region (Bottom)

general for both datasets, especially in the low FPR region. Also, using only position papers does not seem to have a significant negative effect on performance in the low FPR region.

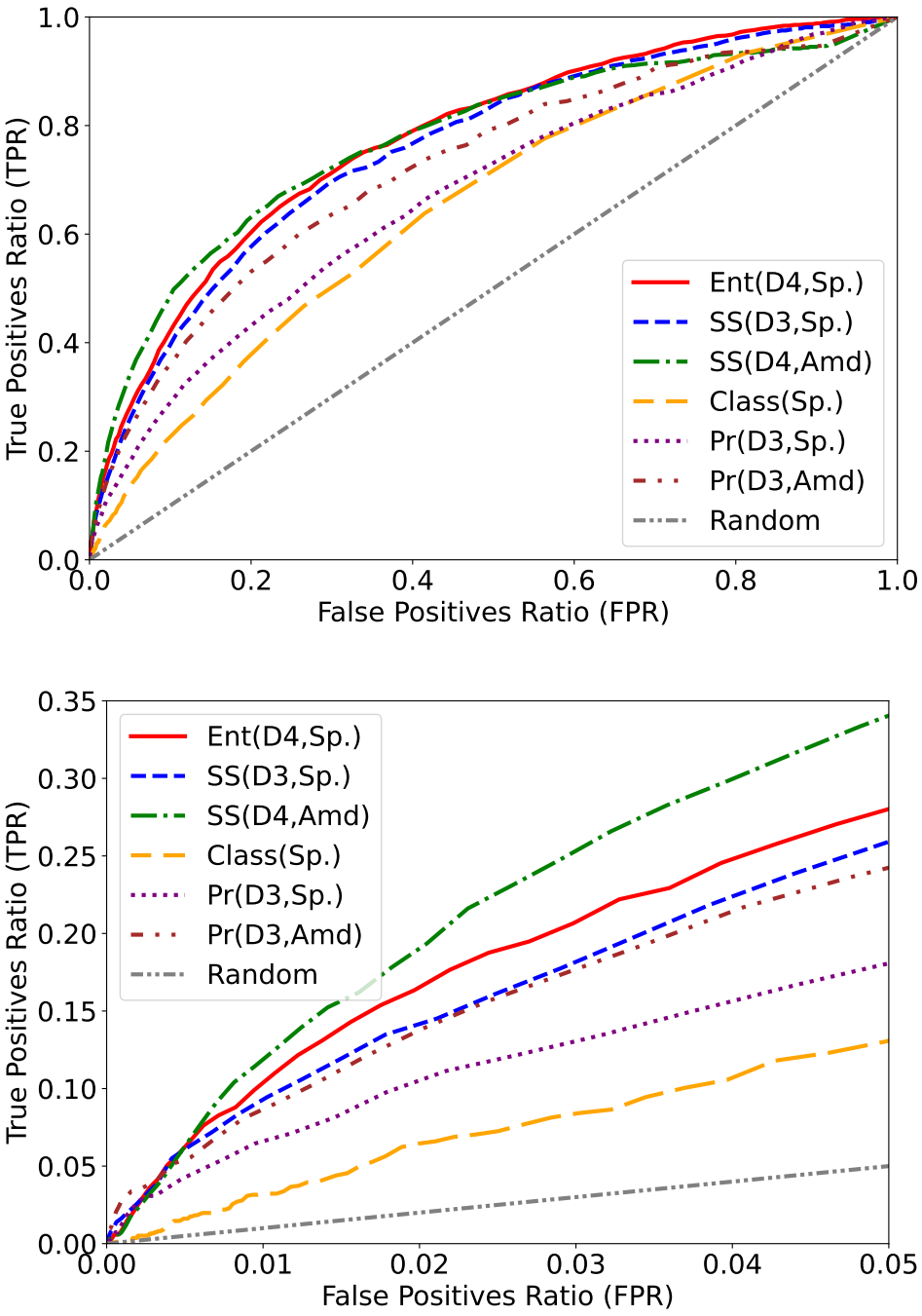

Figure 4.3: ROC curves for the Meetings dataset - Full (Top) and FPR≤0.05 region (Bottom)

It is interesting that the model using amendments performs the best for the meetings, whereas it is worse than the model using speeches in the case of retweets. This could be because amendments are relatively less accessible to the public than speeches, hence this

might reflect links that might not be evident in retweets that are very public. But these links could appear in the meetings data that are relatively less public than retweets.

The entailment method is the best method for the retweet data in the low FPR region and also performs reasonably well in the other cases. Therefore, we use this method for interpretation. Although the improvement over semantic similarity in terms of pAUC is small, entailment significantly improves interpretability by reducing false positive matches in the document pairs, as we show in Section 4.5.3.

4.5 Interpretation

We now interpret the links discovered using the entailment method to see if we can find interesting patterns. To obtain the discovered links, we set the threshold on $A(m, l)$ to 0.7, which gives an FPR of 5% and TPR of 32.5% on the Retweets data. We also manually check a small sample of matched texts and verify that the threshold indeed gives reasonable matches with only a few false positives.

We look at the issues that lobbies are mostly interested in, at their level of focus toward different political groups and ideologies, and at some examples of matched texts that show the method's interpretability.

4.5.1 Lobbies and Debates

We first look at the issues that lobbies are most interested in by ranking the debates, based on the number of links to lobby clusters and after normalizing by the number of speeches in the debate. The list of top-five and bottom-five debates is given in Table 4.6.

Table 4.6: Most and least lobbied debates

Most Lobbied Debates
European Accessibility Act
Packaging and packaging waste, WEEE
Energy efficiency
Plastics in a circular economy
Circular Economy package
Least Lobbied Debates
ESIF: specific measures for Greece
Death penalty in Indonesia
EU-Australia Framework Agreement
Situation in Iraq
Envisaged EU-Mexico PNR agreement

We see that the most lobbied debates are related to energy efficiency and environmental issues (particularly plastic waste, recycling, and the circular economy), whereas the least lobbied debates are related to international agreements and humanitarian issues. The apparent lack of lobbying on international issues could be due to the fact that the governments of countries outside the EU are not required to be registered in the EU TR, hence their lobbying activity is not included in our data.

Similarly, we look at the top debates for specific lobby clusters, and these debates make intuitive sense given the area of interest of the lobby. We give in Table 4.7 the top debates for the *Manufacturing* lobby cluster.

Table 4.7: Top debates for the *Manufacturing* cluster

EU-Korea Free Trade Agreement
European Defence Industrial Development
Anti-dumping, EU steel industry
Common Commercial Policy
Foreign investments in strategic sectors

4.5.2 Lobbies and Political Groups

To evaluate the level of focus for a lobby l towards a particular political group p, we calculate the *lobby focus score*

$$f(l,p) = \frac{n(l,p)}{m_p}, \tag{4.5}$$

where $n(l,p)$ is number of discovered links between l and MEPs in p, and m_p is the number of MEPs in p. To have comparable scores independent of the size of the lobby, we further normalize them as $\hat{f}(l,p) = \frac{f(l,p)}{\max_{p\in\mathcal{P}} f(l,p)}$ where $\mathcal{P}$ is the set of all 9 political groups. We analyze at the level of lobby clusters by averaging $\hat{f}(l,p)$ for all the lobbies l in a particular cluster.

A lobby focus heatmap for selected lobby clusters is given in Figure 4.4. The political groups are ordered in terms of ideology from left to right. We see that lobbies associated with social causes and the environment focus on left-leaning groups, whereas agriculture, ICT, and pharmaceutical lobbies focus more on right-leaning groups.

We also show, in Table 4.8, the left-most and right-most lobby clusters, in terms of the weighted average ideology score of the political groups and with the lobby focus score as the weights. Again, we see that the social and environmental lobbies are aligned to the left, whereas technology, agriculture, and chemical lobbies are aligned to the right.

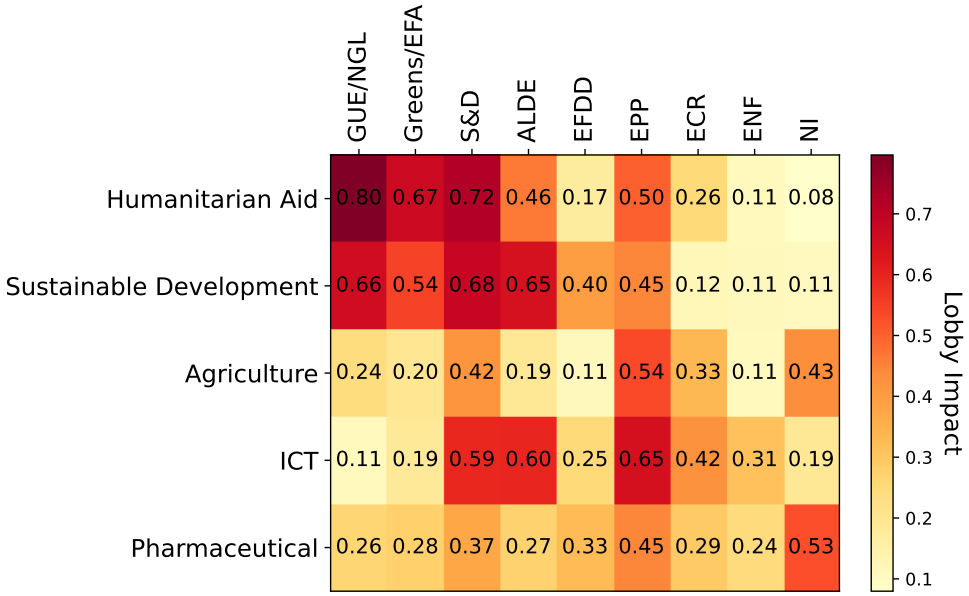

Figure 4.4: Lobby focus heatmap. Political groups are ordered by ideology from left to right.

We construct *focus vectors* for the lobbies

$$\mathbf{f_l} = \left[\hat{f}(l,p) \quad \forall p \in \mathcal{P}\right], \tag{4.6}$$

and obtain the focus vectors for lobby clusters by averaging $\mathbf{f_l}$ for the lobbies l in a cluster. To study how the lobby clusters are arranged in this space, we project them using Principal Component Analysis (PCA).

To interpret each principal component, we compute its Spearman correlation, with the four different ideology scores from the CHES dataset. Only the first three principal components have statistically significant correlations (p-value below 0.0001). The results for these are given in Table 4.9.

We see that PC 3 and PC 2 have strong correlations with general left-right ideology and the economic aspect of ideology respectively. PC 3 also has a strong correlation with the social aspect of ideology.

To visualize and better understand the lobby clusters, in terms of these ideological dimensions, we project them onto PC 2 and PC 3 and obtain the plot in Figure 4.5.

We annotate the dots corresponding to the clusters mentioned in Table 4.8. In addition

Table 4.8: Top and bottom lobby clusters by ideology score. The numbers in parentheses correspond to the numbers in Figure 4.5.

Left-Most Lobby Clusters
Social Economic Interests - 76 (1)
Humanitarian Aid Groups - 19 (2)
Sustainable Development Groups - 12 (3)
HIV/AIDS advocacy and support - 32 (4)
Road safety and transportation advocacy - 66 (5)
Right-Most Lobby Clusters
Technology advocacy groups - 14 (6)
Agricultural interest groups - 64 (7)
Digital and ICT interest groups - 65 (8)
Pharmaceutical and Chemical Advocacy - 48 (9)
Miscellaneous Technology and Education - 79 (10)

Table 4.9: Spearman correlation of principal components with ideology scores. The values in bold have a p-value below 0.0001. The highest absolute values in each row are marked by asterisk(*).

	Ideo	**Econ**	**Soc**	**EU**
PC 1	-0.18	**-0.41**	-0.11	**-0.47***
PC 2	-0.15	**-0.67***	0.02	**-0.47**
PC 3	**0.92***	**0.51**	**0.91**	**-0.44**

to the general left-right placement of these clusters that we already discussed, we also observe their positions with regard to the management of the economy being reflected in the PC 2 coordinates. In particular, the agriculture lobby (number 7) appears to be in favor of more state control (they are known to be in favor of state subsidies (Bednáriková & Jílková, 2012)), whereas the technology lobbies (numbers 6 and 8) appear to advocate for more freedom of the market.

4.5.3 Example Matches

We first look at an example pair of a speech summary and position-paper summary that matched (high semantic similarity and $P(ent) > P(con)$) in Table 4.10. We see clearly that both documents argue in favor of implementing the Pan-European Pension Product (PEPP) and giving it tax advantages at the national level. To demonstrate the advantage of the entailment method, we also show an example pair of a speech summary and position-paper summary that contradict each other (high semantic similarity and $P(con) > P(ent)$) in Table 4.11. We see that though the speech argues in favor of the EU-US Privacy Shield, the position paper opposes it. The entailment method is able to

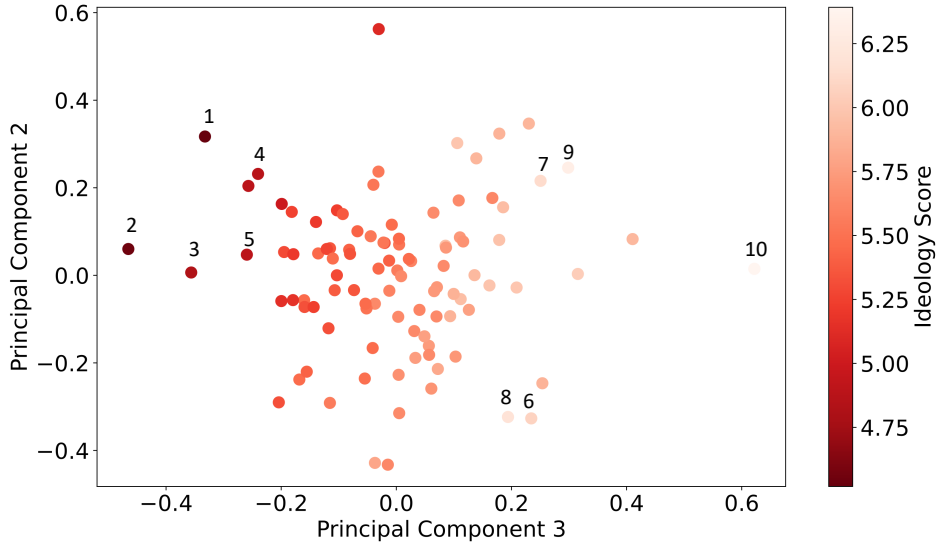

Figure 4.5: Lobby clusters projected on principal components. The color of the dot corresponds to the general left-right ideology score. The dots annotated with numbers correspond to the clusters in Table 4.8.

avoid such false positives.

4.6 Summary

In this chapter, we presented an NLP-based approach for discovering interpretable links between MEPs and lobbies, and we collected novel datasets of position papers, speeches, amendments, tweets, and meetings in the process. We discovered links that were validated indirectly by using tweets and meetings. An aggregate qualitative analysis of discovered links follows expected lines of ideology and the discovered text matches are interpretable. We believe our work will help political scientists, journalists, and transparency activists to have a more efficient and larger-scale investigation of the complex links between interest groups and elected representatives.

Data Limitations The Transparency Register is voluntary for several categories of lobby groups, including public authorities of third countries. We could not also include individual companies that are not part of associations, as position papers are difficult to obtain for them.

Methodology Limitations We considered only English-language lobby documents. There could be some loss of information in the automatic translation of speeches. We could summarise only a limited number of lobby documents due to the cost constraints

Table 4.10: Example matching pair of speech summary s and position paper summary d. Similar portions of the text are highlighted in bold. $\mathbf{v_s}^T\mathbf{v_d} = 0.916$, $P_{(s,d)}(ent) > P_{(s,d)}(con)$.

Speech Summary
We fully **support the implementation of the Pan-European Personal Pension Product (PEPP)** as a means of addressing pension gaps and enhancing cross-border competition in the personal pension market. The PEPP will provide a strong framework for personal pensions with consumer protection and offer new alternatives to those who voluntarily wish to use this scheme. We urge Member States to **grant PEPPs the same tax advantages as similar national products**, and believe that the PEPP is an important step towards building a true pan-European market for personal pension products.
Position Paper Summary
As an interest group operating in the European Parliament, we believe that the **Pan-European Personal Pension Product (PEPP) presents an opportunity** to provide a simpler, more transparent, and cost-effective personal pension solution for EU citizens. The PEPP can increase the mobility of workers in the EU, create a single market for personal pensions and has the potential to boost Europe's capital markets. However, there are many teething problems that need addressing, such as ensuring appropriate regulatory frameworks, making the PEPP simple and transparent, and **addressing national tax incentives**. Ultimately, making the PEPP a mass-market product remains challenging, and **tax incentives are crucial to achieve this goal.**

of using GPT. Open-source LLMs like LLaMA might be useful in this regard.

Release of Data All data is collected from publically available sources. We release data to enable reproducibility while respecting copyrights. The speeches of the MEPs are made publically available by the EP, and their use and reproduction are authorized. For lobby documents, we do not release copies of the original documents. We release only the GPT-generated summaries and the URLs of the original documents. To mitigate link rot, we also release, where possible, links to the archived versions of the documents on the Internet Archive. We ensure that the summaries of position papers that we release do not contain any personal data. Twitter data is collected through their official API and, following their terms of service, we release only the tweet IDs and not the content or metadata of the tweet.

Interpretation of Results We do not claim that the presence of a discovered link between a particular MEP and lobby group necessarily implies that the MEP was influenced, duly or unduly, by the lobby. Rather, it means that the views of the MEP

Table 4.11: Example contradicting pair of speech summary s and position paper summary d. Contradicting portions of the text are highlighted in bold. $\mathbf{v_s}^T\mathbf{v_d} = 0.904$, $P_{(s,d)}(con) > P_{(s,d)}(ent)$.

Speech Summary
We strongly support the importance of transatlantic data transmission for our economy, security, and trade. **The Privacy Shield is a significant step towards achieving much-needed data protection for EU citizens,** and we urge the European Commission to ensure the highest possible standard to avoid legal uncertainty for our companies and SMEs. **It is crucial to have an operational Privacy Shield as soon as possible** for the benefit of our companies, the European economy, and the privacy of EU citizens.
Position Paper Summary
As an interest group operating in the European Parliament, **we have serious concerns about the proposed EU-U.S. Privacy Shield,** which aims to replace the Safe Harbour framework for commercial data flows between the EU and the U.S. **We are urging the European Commission not to adopt the Privacy Shield,** as it does not provide adequate protection for consumers' fundamental rights to privacy and data protection, and fails to address issues related to government surveillance and consumer privacy. We believe that a sustainable arrangement must be established that guarantees privacy protection and legal certainty, based on necessary changes in both the EU and US.

and lobby are probably similar on an issue that is referenced by the matched texts. This similarity could possibly be the result of influence, but such a claim needs to be validated further by the user by a careful examination of the matched texts for their similarity and other relevant information. In this chapter, to avoid any harm to the reputation of MEPs through showing the spurious links that could result from inadvertent errors in our interpretation, we restricted ourselves to an aggregate analysis instead of showing individual MEP-lobby links.

5 Law-Making

In this chapter[1], we study the law-making process *within* parliaments, taking the case of the European Parliament (EP) as an example.

We curate a rich dataset of law amendments proposed by MEPs and develop interpretable models that predict their acceptance by parliamentary committees. Each amendment consists of one or several edits. Edits can be in conflict with edits proposed by other MEPs and with the original proposition in the law. Our models combine three different categories of features: (a) *Explicit* features extracted from data related to the edits, the parliamentarians, and the laws, (b) *latent* features that capture bi-linear interactions between parliamentarians and laws, and (c) *text* features of the edits and laws. We show experimentally that this combination enables us to accurately predict the success of the edits. Furthermore, it leads to model parameters that are interpretable, thus providing valuable insight into the law-making process.

5.1 Introduction

The work of parliaments is governed by complex rules, processes, and conventions, in order to foster compromises among competing viewpoints and priorities. The degree of effectiveness of this process, and the extent it is subject to biases and to benign or undue influences are of obvious concern to both citizens and scientists. An exciting recent development in this regard is the adoption of *open government* principles in many countries (European Union, 2021; Obama White House, 2018; Swiss Government, 2021); the aim is to improve the transparency of the law-making process and the accountability of its protagonists. The EU has been a pioneer in this: It publishes detailed records of

[1]This chapter is based on Kristof et al. (2021). The author of this thesis designed, implemented, and interpreted the part of the model incorporating text features, and collaborated with Dr. Kristof on other parts of the project, including exploratory data analysis, design of baseline models, design and implementation of evaluation experiments for the overall model, error analysis, and solving the cold start problem.

the process by which bills are written and amended until they finally become law. Once an initial draft of a new law has been published, the MEPs in one or several specialized committees examine the draft and propose amendments, consisting of one or more edits. Several edits can be in conflict if they attempt to modify the same part of the draft of the law. To be incorporated into the final version of the law, an edit needs to be approved by the committee in charge, and ultimately by the full plenary of the EP. The EP publishes every proposed edit and its authorship, along with various other details. This makes it possible to build detailed models of the interplay between the MEPs, drafts and edits of the law, and the committees.

Kristof et al. (2020) curate a large-scale dataset of edits proposed by MEPs, over two legislature periods (2009–2019), and they develop a predictive model for the acceptance of proposed edits. They learn a supervised model that endows each law with an "inertia" parameter that captures the difficulty to amend that law, and it endows each MEP with a "strength" parameter that captures the influence or political skill of the MEP. They show that the model achieves good performance, despite its parsimonious structure; in particular, their model does not incorporate any *features* of the laws or edits. This implies an important limitation: Learning the inertia of a law requires training examples of edits success or failure *for that law*. Therefore, we were unable to make a prediction for a new draft of a law for which no edits are contained in the training set[2].

In this chapter, we complement Kristof et al.'s dataset with additional features[3]. Specifically, we collect explicit features for each MEP, including their party membership, country of origin, and gender. We also collect explicit features of the dossiers (law drafts) and edits, including the specific committee in charge and its type. We also collect the actual text of the edits, which enables us to build richer models that take into account the content of the law, as well as the changes affected by each edit. The combination of these explicit features (metadata) and of the text gives rise to models with improved predictive performance. Also, it enables us to make predictions for unseen laws. Finally, we also endow our model with a set of latent features for both laws and MEPs, which capture interactions between MEPs and laws richer than the model in Kristof et al. (2020). Indeed, it is plausible that an MEP be an expert in one subject matter, but less knowledgeable in another; this would bear upon their effectiveness in promoting a particular edit.

Let us briefly summarize our results. We learn a model to predict the adoption or rejection of a proposed edit. An edit can fail because it is rejected in favor of the existing version of the law (the status quo), or because another edit that it is in conflict with is accepted. In our experiments, we report the cross-entropy loss of our predictions. The main results we report assume the *new edit setting* (similar to Kristof et al., 2020), where the edits in the data are randomly split into training, validation, and test sets.

[2]This is reminiscent of the *cold-start problem* in recommender systems.

[3]Data and code publicly available on https://github.com/indy-lab/war-of-words-2

Consequently, most laws that appear in the validation and test sets have edits in the training set. We show that enriching our basic model with additional features results in a significant performance gain, and we explore their relative contributions. This exposes some rather subtle intricacies of the EP's organization and decision-making; for example, we show that the type of committee and the part of the law involved affect the probability of adoption. We also explore how the latent dossier features, learned by the model, cluster into interpretable topics, and we provide some interpretation of the most predictive words and bigrams in an edit. Finally, we apply the model to the more challenging *new law setting*, where a law at test time has not been seen at training time. We show that the features of the text, MEPs, and the dossier can have a predictive value, although the performance is, not surprisingly, lower than in the new edit setting.

The remainder of the chapter is structured as follows. In Section 5.2, we state the problem and provide a detailed description of our dataset. We describe our statistical models in Section 5.3. We give the results and interpretations of our experiments in Section 5.4. We describe related work in Section 5.5 and conclude in Section 5.6.

5.2 Dataset & Problem Statement

5.2.1 The EU Law-Making Process

The legislative process of the EU shares various features with those of liberal democracies. Most laws are created through the *ordinary legislative procedure*, which works as follows. First, the European Commission (*i.e.*, the executive branch of the EU) drafts a law proposal and sends it to the EP (*i.e.*, the representatives of the people in the EU). The EP dispatches the proposal to one of its committees (*e.g.*, for the Agriculture, for Research and Innovation, and for the Economy), whose theme is most closely related to that of the proposal. A committee is a subset of the MEPs. For example, if the proposal is about limiting carbon emissions in the EU, it will go to the Environment Committee.

One MEP in the committee is elected as the *rapporteur*, *i.e.*, as the person in charge of the proposal for the committee. The rapporteur and all other MEPs in the committee can propose *amendments* to the proposal, *i.e.*, modifications to parts of the law. An amendment consists of one or several *edits*, *i.e.*, a sequence of contiguous words that are added to or removed from the proposal text. These edits can conflict with other edits if they attempt to change the same part of the law but in different ways. The members of the committee vote on each edit to decide whether to include it or not in the final *report*; this decision forms the position of the Parliament on the proposal. The report is then voted on by the whole Parliament: If it is accepted, it is transferred to the Council of Ministers (*i.e.*, the equivalent of a senate representing the member states of the EU). If it is rejected, the proposal is abandoned.

Amendment 802
Lidia Joanna Geringer de Oedenberg, Catherine Stihler, Victor Negrescu
Article 13 – title

Text proposed by the Commission	*Amendment*
Use of protected content ***by*** information society service providers ***storing and giving access to large amounts of works and other subject-matter uploaded by their users***	Use of ***copyright*** protected content ***uploaded by users of*** information society service providers

Amendment 803
Tadeusz Zwiefka, Bogdan Brunon Wenta
Article 13 – title

Text proposed by the Commission	*Amendment*
Use of protected content by information society service providers storing and giving access to ***large*** amounts of works and other subject-matter uploaded by their users	Use of protected content by information society service providers storing and giving access to ***significant*** amounts of ***copyright protected*** works and other subject-matter uploaded by their users

Figure 5.1: Example of two conflicting amendments in their raw format on the title of Article 13 of a proposal about copyright on the Internet. (Top) Am. 802 is proposed by three MEPs and consists of three edits. (Bottom) Am. 803 is proposed by two other MEPs on the same text, and it consists of two edits. The last edit of Am. 802 (deleting the end of the title) conflicts with both edits of Am. 803. Only the first edit of Am. 803 (replacing "large" by "significant") was accepted, and all other edits were rejected.

Optionally, the MEPs in another committee can decide that their expertise is relevant to the proposal. For example, the Transportation Committee could also want to make amendments to the proposal about limiting carbon emissions in the EU. Hence, they can send their *opinion* to the reporting committee, *i.e.*, their suggested amendments (edits) to the proposal. The process is similar to that of creating a report: A rapporteur is elected to be in charge of the opinion and can, together with other MEPs in the opinion committee, propose amendments. The opinion differs from the report in that it is not voted by the whole Parliament (only the report is), and the reporting committee is free to take into account the amendments from the opinion. Amendments from the opinion committee can be in conflict, however, with amendments from the reporting committee, and the MEPs from the reporting committee will also have to vote on those. Using the existing terminology, we will refer to reports and opinions as *dossiers*. A more detailed description of the European legislative process can be found in Kristof et al. (2020).

We show an example of conflicting edits in two amendments in Figure 5.1. The two amendments are proposed in Article 13 of a proposal about copyrights on the Internet. Amendment 802 is proposed by three MEPs and consists of three edits: (a) Inserting "copyright" (in green), (b) replacing "by" by "uploaded by users of" (in yellow), and (c)

deleting the end of the title after "providers" (in red). Amendment 803 is proposed by two other MEPs and consists of two edits: (d) Replacing "large" by "significant" (in yellow) and (e) inserting "copyright protected" (in green). There are two conflicts in this amendment: Edit (c) of the first amendment is in conflict with Edit (d), and it is also in conflict with Edit (e). All these edits are also implicitly in conflict with the original text proposed by the European Commission. Of these five edits, only Edit (d) was accepted. All other edits were rejected, *i.e.*, the status quo was voted and the text proposed by the Commission was maintained.

5.2.2 Explicit Features

The dataset in Kristof et al. (2020) contained the following metadata: (a) The author(s) of an amendment (b), the dossier that is amended, and (c) the rapporteur for this dossier. We complement this dataset by extracting explicit (meta) features of the MEPs, the edits, and the dossiers, as well as text features. For each MEP, we collect information on their nationality (one of 28), their EU political group (one of 9), and their gender. A political group consists of national parties that share similar political ideologies. For each edit, we identify whether it is an insertion, a deletion, or a replacement of some words in the proposal, and we compute its length. We also collect information about where in the law the edit was proposed: in an article (in the body of the proposal), in a recital (in the preamble of the proposal), in an annex, or in other more specific but less frequent parts of a law. We determine whether an edit in a reporting committee comes from an opinion committee (in which case it is an "outsider"). Finally, we note whether an edit comes with an optional justification. For each dossier, we identify its type (report or opinion) and the committee that is in charge. We also note if the proposal is a regulation (legally binding for all member states of the EU), a directive (sets general goals that member states can implement however they want), or a decision (binding to one member state or company only). We describe these explicit features in Table 5.1.

In total, we collect 449,493 edits from 237,177 amendments in the European Parliament during the seventh and the eighth legislature periods (referred to as EP7 and EP8), between 2009 and 2019 (each period lasts 5 years). After gathering the edits according to the conflicts, we obtain 267,451 conflicts for both EP7 and EP8, covering 1,889 dossiers. We summarize this dataset in Table 5.2.

5.2.3 Text Features

We further augment the dataset by collecting text features of the edit itself. It is reasonable to expect that certain words and phrases are predictive of the success of an edit. We extract the deleted words w_- from the proposal and the inserted words w_+ from the amendment. In Figure 5.1, for example, Edit (b) of Amendment 802 has w_- = *"by"*

Table 5.1: List of features for MEPs and edits.

Category	Feature	Type [Values]
MEP	Nationality	Categorical [28]
	Political group	Categorical [8 or 9]
	Gender	Categorical [2]
	Rapporteur	Binary
Edit	Edit type	Categorical [3]
	Log-length (+)	Numerical [$\mathbb{R}_{\geq 0}$]
	Log-length (-)	Numerical [$\mathbb{R}_{\geq 0}$]
	Article type	Categorical [7]
	Outsider committee	Binary
	Justification	Binary
Dossier	Type	Categorical [2]
	Committee	Categorical [35]
	Legal act	Categorical [3]

Table 5.2: Dataset Statistics

	EP7 (2009–2014)	**EP8 (2014–2019)**
Number of Amendments	108,292	128,885
Number of Edits	200,407	249,086
Number of Conflicts	126,417	141,034
Number of MEPs	761	791
Number of Dossiers	1,089	800
Edits Accepted (%)	37.7	25.7
Insertions (%)	37.8	37.9
Deletions (%)	22.0	22.4
Replacements (%)	40.2	39.7

and w_+ = *"uploaded by users of"*. We also consider the context of an edit by extracting the original text of the amended article, surrounding the location of the edit. For Edit (b) in Amendment 802, the context consists of the two portions of text *"Use of protected content"* and *"information society...their users"*. Finally, we also extract the title of the law proposal; we will use it as a text feature of the dossier. For Amendments 802 and 803, the title is *"Copyright in the Digital Single Market"*. We map all words to lowercase, and we replace digits in the title with the letter "D", as there are many reference numbers that are unlikely to be useful for our task.

We give some statistics of the distribution of the length of the deleted text w_-, the inserted text w_+, the context, and the title in Table 5.3. We report the lower quartile Q_1 and the upper quartile Q_3, as well as the median. Approximately half of the inserted and

Table 5.3: Distribution of text lengths (in numbers of words).

Legislature	Type	Q_1	Median	Q_3
EP7	Insertion w_+	2	7	20
	Deletion w_-	2	6	26
	Context	15	42	79
	Title	6	12	19
EP8	Insertion w_+	2	6	17
	Deletion w_-	2	6	28
	Context	20	49	93
	Title	6	10	22

deleted texts are short (7 words or less), but the distribution of lengths has a long tail, as shown by the larger values of the upper quartile Q_3. The context provides large portions of the text (the median is at 42 for EP7 and 49 for EP8), which will be useful for making predictions. In Section 5.3, we describe how we incorporate the explicit features and the text features into our models.

5.2.4 Problem Statement

We build a model that predicts the vote outcome of edits that will form the reports and the opinions. Formally, we take a supervised approach to solve the following prediction problem: Let $\mathcal{C} = \{a, b, \ldots\}$ be a set of conflicting edits proposed on a dossier i, for which we have observed other edits. We want to predict which of the conflicting edits in $\mathcal{C}$ or the status quo of the proposal for dossier i will be accepted by the committee. This task differs from multinomial classification as the number of classes varies for each data point: If an edit a is in conflict only with the original text proposed by the Commission, then $|\mathcal{C}| = 1$. If several edits $a, b, \ldots \in \mathcal{C}$ are in conflict with each other, then $|\mathcal{C}| > 1$.

According to Rule 180 of the Rules of Procedure of the European Parliament (European Parliament, 2021b), the committee sets a deadline by which the MEPs must propose amendments to a dossier. The voting takes place after this time. Hence, at the time of voting, an edit is expected to confront all the alternatives: If edits a, b, and c are in conflict, the MEPs vote on all three of them and on the status quo in order to select only one outcome.

5.3 Models

To better introduce our models, we first define the baselines against which we will compare our results in Section 5.4. In particular, we recall the War of Words model, as introduced in Kristof et al. (2020), and we adopt the same terminology for consistency.

For each baseline and for our models, we assume a set of K conflicting edits $\mathcal{C} = \{a, b, \ldots\}$ proposed on dossier i, for which we want to model the probability that an edit $a \in \mathcal{C}$ is accepted over edits $b, \ldots$ on this dossier. We denote this probability by $P\left(a \succ_i \mathcal{C} - \{a\}\right)$, and we denote the probability that the status quo wins, *i.e.*, that the original text proposed by the Commission is kept, by $P\left(i \succ \mathcal{C}\right) = 1 - \sum_{a \in \mathcal{C}} P\left(a \succ_i \mathcal{C} - \{a\}\right)$.

5.3.1 Baselines

Naive Classifier The *naive classifier* predicts a uniform probability for each outcome, *i.e.*, for each of the conflicting edits or the status quo to win, as

$$p\left(a \succ_i \mathcal{C} - \{a\}\right) = p\left(i \succ \mathcal{C}\right) = \frac{1}{K+1}.$$

Random Classifier The *random classifier* learns the prior probability $p^{(K)}$ that the status quo wins for each conflict size $|\mathcal{C}| = K$, and it predicts

$$p\left(i \succ \mathcal{C}\right) = p^{(K)}.$$

It predicts uniformly each of the edits to win as

$$p\left(a \succ_i \mathcal{C} - \{a\}\right) = \frac{1 - p^{(K)}}{K}.$$

War of Words The WoW model encodes (a) the collaboration between MEPs who co-sponsor an edit and (b) the conflicts between edits as a discrete-choice model reminiscent of the Bradley-Terry model (Bradley & Terry, 1952).

It models the probability that an edit a is accepted over edits $b, \ldots$ on dossier i as

$$p\left(a \succ_i \mathcal{C} - \{a\}\right) = \frac{\exp(s_a)}{\sum\limits_{c \in \mathcal{C}} \exp(s_c) + \exp(d_i + b)}, \tag{5.1}$$

where $s_a = \sum_{u \in \mathcal{A}_a} s_u$ is the cumulated *skill* of all authors $\mathcal{A}_a$ of edit a, $d_i \in \mathbb{R}$ is the *difficulty* of dossier i, and $b \in \mathbb{R}$ is a global bias parameter. The skill parameters s_u of the MEPs can be interpreted as a measure of their influence, and the difficulty parameters d_i of the dossiers can be interpreted as a measure of their controversy.

5.3.2 Enriched Models

Explicit Features We extend the WoW model by augmenting it with explicit features of the MEPs (e.g., nationality), the edits (e.g., length of inserted text), and the dossiers

(e.g., report or opinion), as described in Table 5.1. From (5.1), we replace the skill parameters s_a with the inner product between a feature vector $\mathbf{s_a} \in \mathbb{R}^{M_E}$ of M_E features of edit a and the associated parameter vector $\mathbf{w_E} \in \mathbb{R}^{M_E}$. We also replace the difficulty parameter d_i by the product of a feature vector $\mathbf{d_i} \in \mathbb{R}^{M_D}$ of M_D features of dossier i and its associated parameter vector $\mathbf{w_D} \in \mathbb{R}^{M_D}$. We then have

$$P\left(a \succ_i \mathcal{C} - \{a\}\right) = \frac{\exp(\mathbf{s_a}^T \mathbf{w_E})}{\sum\limits_{c \in \mathcal{C}} \exp(\mathbf{s_c}^T \mathbf{w_E}) + \exp(\mathbf{d_i}^T \mathbf{w_D} + b)}. \tag{5.2}$$

We refer to this model as WOW(*Explicit*) (or WOW(X), for conciseness). In (5.1), the feature vector $\mathbf{s_a}$ is the indicator of the authors of an edit a: Its entries s_u are 1 for all $u \in \mathcal{A}_a$ and 0 otherwise. Similarly, the feature vector $\mathbf{d_i}$ is the indicator of dossier i. In (5.2), the feature vectors $\mathbf{s_a}$ and $\mathbf{d_i}$ represent features related to the MEPs, the edits, and the dossiers derived from our dataset.

Latent Features Consider the simple case of an MEP u proposing an edit on dossier i, and suppose that this edit conflicts with another edit, proposed by MEP v. From (5.1), let $P(u \succ_i v)$ be the probability that, for dossier i, the edit proposed by MEP u is accepted over the edit proposed by MEP v. The assumption made in the WOW model is strong: It posits that if MEP u is more influential than MEP v, then, all other parameters being equal, $P(u \succ_i v) > P(v \succ_i u)$ for all dossiers i. This assumption is not always realistic: The dossiers span a vast amount of different topics, and the MEPs have their own specializations and interests. For example, an MEP familiar with fisheries might not be knowledgeable about research and academia.

In order to capture these dependencies, we incorporate a bi-linear term into the WOW model. We assign a vector $\mathbf{x_u} \in \mathbb{R}^L$ to each MEP u, and a vector $\mathbf{y_i} \in \mathbb{R}^L$ to each dossier i, for some dimensionality $L > 0$. We then rewrite (5.1) as

$$P\left(a \succ_i \mathcal{C} - \{a\}\right) = \frac{\exp(s_a + \mathbf{x_a}^T \mathbf{y_i})}{\sum\limits_{c \in \mathcal{C}} \exp(s_c + \mathbf{x_c}^T \mathbf{y_i}) + \exp(d_i + b)}, \tag{5.3}$$

where $\mathbf{x_a} = \sum_{u \in \mathcal{A}_a} \mathbf{x_u}$ is the sum of the latent features $\mathbf{x_u}$ of each author u of edit a. We refer to this model as the WOW(*Latent*) model (or WOW(L)). The latent vectors $\mathbf{x_u}$ and $\mathbf{y_i}$ can be viewed as the embeddings of MEP u and of dossier i in a Euclidean latent space. Informally, the probability $P\left(a \succ_i \mathcal{C} - \{a\}\right)$ increases when the MEP embedding $\mathbf{x_a}$ is co-linear with the dossier embedding $\mathbf{y_i}$ in the latent space. It decreases when the two vectors point in opposite directions. Furthermore, vector $\mathbf{x_u}$ can be interpreted as the set of skills of MEP u. Similarly, $\mathbf{y_i}$ can be interpreted as the set of skills required to edit dossier i.

Text Features The features described so far ignore the text content of the edit itself. It is reasonable to expect that the presence of certain words or phrases in the original or amended text of an edit, and in the title of the dossier, are predictive of the success of the edit. Hence, we incorporate text features to the WOW model by rewriting (5.1) as

$$P\left(a \succ_i \mathcal{C} - \{a\}\right) = \frac{\exp(s_a + \mathbf{r_a}^T \mathbf{w_T})}{\sum_{c \in \mathcal{C}} \exp(s_c + \mathbf{r_c}^T \mathbf{w_T}) + \exp(d_i + \mathbf{r_i}^T \mathbf{w_{T'}} + b)}, \tag{5.4}$$

where $\mathbf{r_a} \in \mathbb{R}^D$, $\mathbf{r_i} \in \mathbb{R}^{D'}$ are, respectively, representations of the text of the edit a and the title of dossier i, and $\mathbf{w_T} \in \mathbb{R}^D$, $\mathbf{w'_T} \in \mathbb{R}^{D'}$ are, respectively, the associated parameter vectors. We refer to this model as the WOW(*Text*) model (or WOW(T)).

We explore two ways of learning the representations $\mathbf{r_a}$ and $\mathbf{r_i}$: (1) from pre-trained word embeddings and (2) by training embeddings on our dataset. With pre-trained embeddings, $\mathbf{r_a}$ is the concatenation of three vectors that are the representations of the deleted text, inserted text, and the context of the edit, as explained in Section 5.2. Each of these vectors is the average of the pre-trained word embeddings of the words in these parts of the text, and $\mathbf{r_i}$ is the average of the pre-trained embeddings of the words in the title of dossier i. We use two sets of pre-trained embeddings trained with the word2vec algorithm (Mikolov et al., 2013): (a) 300-dimensional embeddings trained on Google News (Google, 2013) and (b) 200-dimensional Law2Vec embeddings trained on legal texts of the EU, the US, the UK, Canada, and Japan(Chalkidis & Kampas, 2019).

We also learn embeddings from our dataset by using the fastText (Supervised) model for text classification (Joulin et al., 2017). In the simplest version of this model, a D-dimensional embedding is learned for each word (and n-grams) in a dataset. A piece of text is then classified with a softmax layer by representing it as the average of the word embeddings. To construct $\mathbf{r_a}$ and $\mathbf{r_i}$, we use the learned word and bigram embeddings.

The original fastText model is defined, however, for the classification of homogeneous pieces of text into a fixed set of classes. This does not directly apply to our problem, as (a) the text features for the edit are of three types (deleted text, inserted text, and context) and (b) the size of a conflict $|\mathcal{C}| = K$ varies from a data point to another. We solve the first problem by prepending tags (`<del>`, `<ins>`, and `<con>`) to each word to enable the model to learn separate embeddings for the same word in different types of text feature. We solve the second problem by training the embeddings on a binary classification task of edit acceptance (based only on the text), and by using the embeddings learned on this ad-hoc task into the WOW models. We learn the embeddings for the words in the title by training a different fastText model to predict the acceptance of an edit from the title only. This is equivalent to predicting the probability of acceptance of the status quo for each dossier, given its title. For our experiments in Section 5.4, we use the fastText embeddings rather than pre-trained embeddings, because the former performed better on the ad-hoc binary classification task.

Table 5.4: Variations of our model by combinations of features (explicit, latent, and text features).

Model	Equation	Explicit	Latent	Text
WoW	(5.1)	–	–	–
WoW(*Explicit*)	(5.2)	✓	–	–
WoW(*Latent*)	(5.3)	–	✓	–
WoW(*Text*)	(5.4)	–	–	✓
WoW(*XL*)	(5.2) & (5.3)	✓	✓	–
WoW(*XT*)	(5.2) & (5.4)	✓	–	✓
WoW(*LT*)	(5.3) & (5.4)	–	✓	✓
WoW(*XLT*)	(5.2), (5.3) & (5.4)	✓	✓	✓

Hybrid Models To obtain hybrid models with different components, we combine WoW(*Explicit*), WoW(*Latent*), and WoW(*Text*). This helps us understand the contribution of each type of feature to the performance, in Section 5.4. We summarize all the possible combinations in Table 5.4, and we sort them by increasing levels of complexity. The WoW model has no features at all and will serve as a baseline. The WoW(*XLT*) combines explicit, latent, and text features, and it has the highest complexity.

5.3.3 Learning the Parameters

Each observation n is a triplet $(\mathcal{C}_n, i_n, l_n)$ of (a) a set of conflicting edits $\mathcal{C}_n$ with $|\mathcal{C}_n| = K_n > 0$, (b) a dossier i_n on which the edits are proposed, and (c) a label $l_n \in \mathcal{C}_k \cup \{i_n\}$ indicating which of the K_n edits or the status quo is accepted. We assume that the triplets are independent. Given a dataset of N triplets $\mathcal{D} = \{(\mathcal{C}_n, i_n, l_n) \mid n = 1, ..., N\}$ and given a vector θ of all the parameters in our model, we learn θ by minimizing their negative log-likelihood under $\mathcal{D}$

$$L(\theta; \mathcal{D}) = \sum_{n=1}^{N} \sum_{a \in \mathcal{C}_n} \left[\mathbb{1}_{\{l_n = a\}} \log P\left(a \succ_{i_n} \mathcal{C}_n - \{a\}\right) + \mathbb{1}_{\{l_n = i_n\}} \log P\left(i_n \succ \mathcal{C}_n\right) \right], \quad (5.5)$$

where $P\left(a \succ_{i_k} \mathcal{C}_k - \{a\}\right)$ and $P\left(i_k \succ \mathcal{C}_k\right)$ depend on θ. In order to avoid overfitting, we add regularization to the negative log-likelihood. We pre-process our dataset by keeping only the dossiers for which more than 10 edits are proposed and only the MEPs who proposed more than 10 edits. Thus, we obtain a dataset of $N = 125,733$ data points for EP7 and $N = 140,763$ data points for EP8. In the WoW(*Explicit*) and the WoW(*Text*) models, the log-likelihood is convex, and we find optimal parameters by using an off-the-shelf convex optimizer (L-BFGS-B (Byrd et al., 1995)). In the WoW(*Latent*) model, the bi-linear term breaks the convexity, and we can no longer ensure that we will find parameters that are global optimizers. In practice, by using a stochastic gradient descent algorithm (Adagrad (Duchi et al., 2011)), we are still able to find good model parameters

without convergence issues.

5.4 Results

5.4.1 Experimental Setting

We report the cross-entropy loss to evaluate the baselines and our models. Let $(\mathcal{C}_n, i_n, l_n)$ be an observation. We compute

$$\ell_n = \begin{cases} -\log P(l_n \succ_{i_n} \mathcal{C}_n - \{l_n\}) & \text{if } l_n \in \mathcal{C}_n, \\ -\log P(i_n \succ \mathcal{C}_n) & \text{if } l_n = i_n. \end{cases} \tag{5.6}$$

We report the average value for all N points in our test set as $\ell = \frac{1}{N}\sum_n \ell_n$. We randomize our dataset and split it into 80% for training, 10% for validation, and 10% for the final evaluation. Note that an edit can be involved in several conflicts. For example, in Figure 5.1, edit c is involved in two conflicts: $\mathcal{C}_1 = \{c, d\}$ and $\mathcal{C}_2 = \{c, e\}$. Hence, we assign conflicts to each set so that an edit is present in exactly one set. We combine both the training and the validation sets to fit our model, before evaluating it on the test set. We set the number of latent dimensions L and the regularizers, and we choose the best word embeddings, by held-out validation. This results in fastText of dimension $D = D' = 10$, with bigrams.

5.4.2 Predictive Performance

We show in Figure 5.2 the overall performance of all variations of our model (with and without explicit, latent, and text features) over EP7 and EP8, and we compare them against the naive and the random predictors, as well as against the WoW model. All our models outperform the baselines, and WoW(XLT) outperforms all other models. Including explicit features improves the performance of the predictions in terms of cross entropy by 7% for EP7 and 6% for EP8 over the simpler WoW model. On EP7, WoW(L) improves the performance by 12% and WoW(T) by 7%, whereas for EP8 the difference between the two models is smaller (10% increase for WoW(L) and 8% for WoW(T)). Hence, the text features provide a greater improvement for EP8 than for EP7, whereas the latent features provide a greater improvement for EP7 than for EP8. The difference between WoW(XL) and WoW(L) (0.010 for EP7 and 0.013 for EP8) is less than the difference between WoW(XT) and WoW(T) (0.034 for EP7 and 0.035 for EP8). This suggests that the information in latent features has a significant overlap with that in explicit features, whereas text features provide more complementary information. Finally, combining the text and latent features provides high performance, but combining them further with explicit features leads to the best performance.

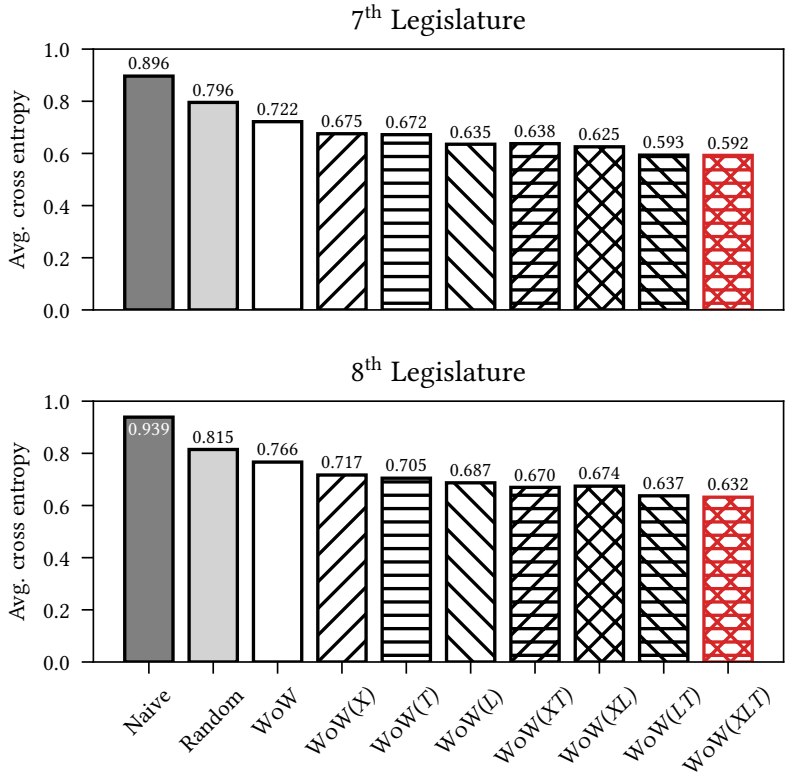

Figure 5.2: Average cross-entropy loss of the baselines and our models. Combining the explicit, latent, and text features help obtain the best performance.

5.4.3 Interpretation of Explicit Features

To understand the contribution of the explicit features to the predictive performance, we show in Figure 5.3 the decrease in cross-entropy loss of WoW(*MEP*) (all MEP features except the rapporteur feature), WoW(*Rapporteur*) (rapporteur feature only), WoW(*Edit*), and WoW(*Dossier*) over WoW. The dossier features contribute virtually nothing to the predictive performance (the difference is at the fourth decimal point). Similarly for EP7, the nationality, political group, and the gender features of WoW(*MEP*) contribute very little. For EP8, these features improve the performance, but not as much as the edit features. This suggests that these features have limited influence on the predictions. The nationalities and political groups are qualitatively analyzed in the literature in the context of their influence on MEPs' voting behaviour (Coman, 2009; Hix, 2002; Lefkofridi & Katsanidou, 2014; Mühlböck, 2012). To the best of our knowledge, there is no analysis of their effect on the amending process. Interestingly, for EP7, combining all features into the WoW(X) model leads to a performance boost that is greater than the sum of each individual feature group.

To gain insight into the dynamics of the legislative process, we interpret the values of the parameters of WoW(XLT) trained on the full dataset for EP8 (combining training,

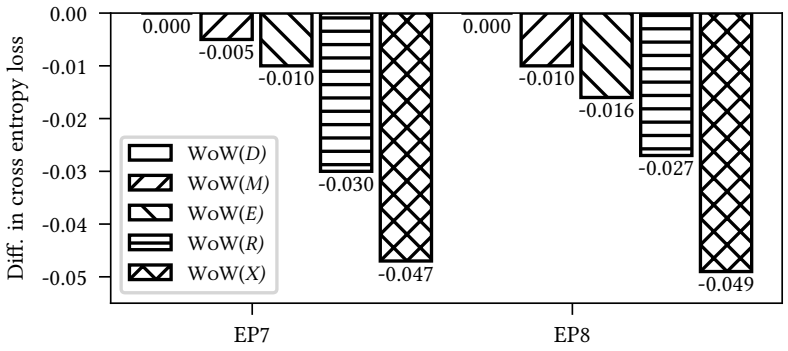

Figure 5.3: The difference in cross-entropy loss over WoW of different models. The rapporteur feature and the edit features contribute more to the predictive performance than the MEP and dossier features.

validation, and test data). Let $w_f \in \mathbb{R}$ be the value of the parameter associated with feature f. The rapporteur feature r of WoW(*Rapp.*) provides a greater decrease in loss. This *rapporteur advantage* complements the findings of Costello and Thomson (2010), conducted by interviewing key informants over EP5 (1999-2004) and EP6 (2004-2009). They show that the rapporteur, with their particular role, has some influence on the legislative process, albeit constrained. We note that, according to our model, the rapporteur advantage has slightly increased in EP8 ($w_r = 1.19$) compared to EP7 ($w_r = 1.12$).

These explicit features enable us to explain the contributions to the success of an edit. We report here (and in subsequent sections) the results for EP8 only. All other parameters being equal, a female ($w_{\text{fem}} = -0.02 > -0.04 = w_{\text{mal}}$) MEP from Latvia, whose party belongs to the group of the European People's Party (center-right), has the highest chance to see her edit accepted. This edit has even higher chances if it inserts ($w_{\text{ins}} = -0.03 > w_{\text{del}} = -0.13 > w_{\text{rep}} = -0.22$) a short portion of text (the feature associated with both insertion and deletion length is negative) in a part of the law that is not its body or its preamble (w_{art}, w_{rec} and w_{para} have the lowest value among the seven article types). Adding a justification also increases the probability of an edit being accepted ($w_{\text{jus}} = 0.08$), as well as edits from the opinion committee (referred to as the "outsider committee" feature in Table 5.1, $w_{\text{out}} = 0.16$).

For the dossier features, our model learns that it is harder to make edits on reports, compared to opinions ($w_{\text{rep}} = 0.33 > -0.26 = w_{\text{opi}}$). As explained in Section 2.3, reports are voted on by the whole Parliament. Therefore, the reports have a greater influence on the final law, and we expect that the MEPs make it more difficult for competing edits to be accepted in reports. Finally, our model also learns that it is harder to make edits for decisions and directives, compared to regulations ($w_{\text{dec}} = 0.25 > w_{\text{dir}} = 0.12 > w_{\text{reg}} = 0.10$).

5.4.4 Interpretation of Text Features

In Figure 5.2, we observe that the text features contribute significantly to improving the performance. We use the learned parameter vectors $\mathbf{w_T}$ and $\mathbf{w_{T'}}$ of WOW(XLT) to identify words and bigrams that have the most predictive power. First, we rank the words and bigrams of the edit text, according to the dot product of their embeddings with $\mathbf{w_T}$. The top-k terms (having a positive dot product) contribute the most towards acceptance of the edit, whereas the bottom-k terms (having a negative dot product) contribute most towards rejection of the edit. The opposite holds for the terms of the title and their dot product with $\mathbf{w_{T'}}$.

We look at the top 50 terms for each feature and prediction outcome, and we find some interesting patterns among these terms, although not all of them are easy to interpret. Note that we have more than 10,000 unique terms for the edit text and more than 1,000 unique terms for the title, hence we consider only the most predictive terms near the ends of the ranking. A list of the top-50 terms for each feature and prediction outcome is reported in Appendix C.

One of the bigrams that, when deleted, is predictive of acceptance is *any other*, which is commonly used to widen the scope of the law (as in "contractual or any other duty"). Interestingly, the bigrams of *human rights* and *data protection* are also predictive of acceptance when deleted. The word *should*, which is used to add recommendations, is predictive of acceptance when inserted, whereas adding *must*, which is used for obligations, is predictive of rejection. We see that *best* is predictive of acceptance, which is commonly used to make a requirement stronger (as in "best available scientific evidence", "best possible way"). Adding *positive* and *positive impact* predicts acceptance, whereas adding *negative* predicts rejection. Adding the word *inserted*, which commonly refers to inserting new articles in existing laws, is predictive of acceptance, whereas *deleted* is predictive of rejection.

Considering the words in the context, we see that *firearms*, *resettlement*, *terrorist*, and *fingerprints* are predictive of rejection. This could be because the laws related to these topics are controversial, hence many edits are rejected due to conflicts. For the words in the title, we see that *customs*, *community*, *financial*, *fisheries*, and *general budget* are predictive of acceptance, whereas *market*, *framework*, *structural reform*, *emission*, and *greenhouse gas* are predictive of rejection. This suggests the relative ease or difficulty of editing laws related to these topics, and it correlates well with the values of the difficulty parameters d_i: The top-50 dossiers with the highest difficulty parameters contain highly controversial dossiers about establishing frameworks for the screening of foreign investments and vast public investment programs (InvestEU and Horizon Europe), as well as regulation of the financial market, copyrights in the digital market, and carbon-emission reduction. The bottom-50 dossiers with the lowest difficulty parameters contain low-controversy dossiers about cohesion within the EU, financial rules, fisheries,

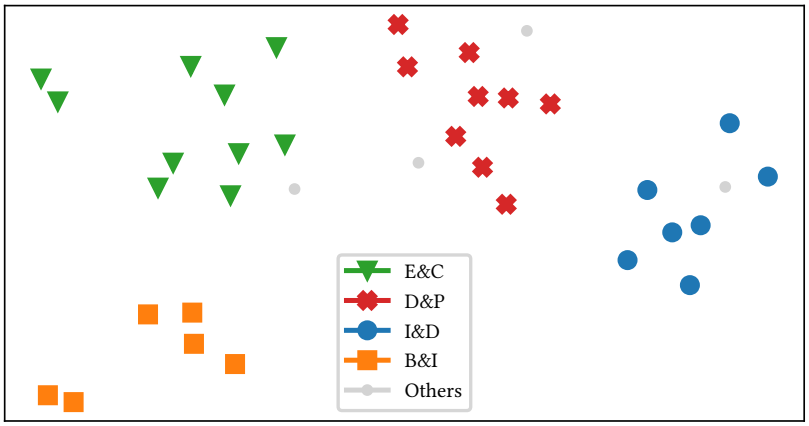

Figure 5.4: Visualization with t-SNE of the top-10 and bottom-10 dossiers on the first two principal components in EP8. There are four clusters: Environment and Communication (E&C), Defense and Protection (D&P), Investment and Development (I&D), and Business and Innovation (B&I).

and the community code on visas.

5.4.5 Interpretation of Latent Features

The latent features improve the predictions overall and help capture the complex dynamics of the legislative process. The best number of latent dimensions is $L = 20$ for the models including latent features. In order to interpret the latent features, we gather the latent vectors $\mathbf{y_i}$ learned by WoW(XLT) into a matrix $\mathbf{Y} = [\mathbf{y_i}]$. We apply principal component analysis and keep the top-10 and bottom-10 dossiers from each of the first two principal components in EP8. We use t-SNE (Maaten & Hinton, 2008) to represent these forty dossiers in a two-dimensional space, and we show the projection in Figure 5.4.

We distinguish four clusters. The cluster at the top-left contains dossiers about fuel quality, renewable energy, trade of animals, and sustainable investments. It also contains dossiers about electronic communications, the processing of personal data, and the sharing of public information. We interpret this cluster as *environment and communication*, and we highlight with green triangles the corresponding dossiers. The cluster at the top-center contains dossiers about the establishment of defense funds, the prosecution of criminal offenses, and the identification of criminals between member states. It also contains dossiers about the protection of workers, businesses, refugees, internal markets, and cultural goods. We interpret this cluster as *defense and protection* (red crosses). The cluster at the top-right contains dossiers about vast investment and development programs, finance, and the development of internal markets. We interpret this cluster as *investment and development* (blue dots). Finally, the cluster at the bottom-left contains dossiers about economic competitiveness and innovation, as well as frameworks for business development and the funding of start-up companies. We interpret this cluster

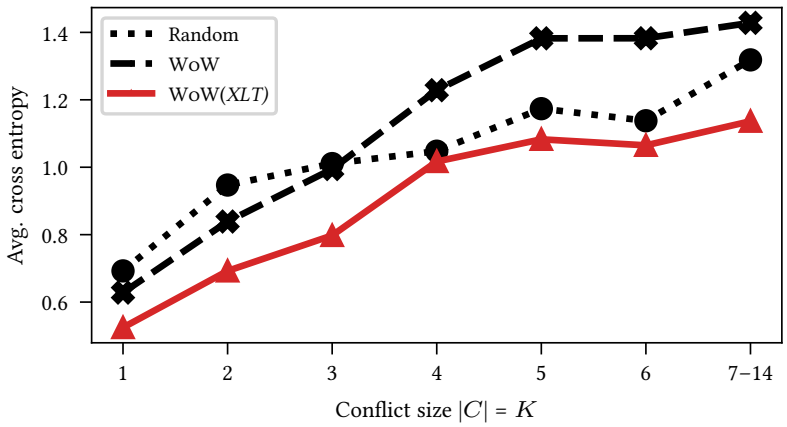

Figure 5.5: Average cross-entropy loss per conflict size $|\mathcal{C}| = K$. The loss of the WoW(XLT) model increases less rapidly than the loss of the baselines.

as *business and innovation* (orange squares).

5.4.6 Error Analysis by Conflict Size

We explore how the WoW(XLT) model performs on conflicts of different sizes in the test set for EP8 (we observe a similar behavior on EP7). We bin the conflict size so that there are at least 100 data points in each bin. The distribution of conflict size is exponentially decreasing: There are 8,462 conflicts of size 1 (i.e., an edit is in conflict with the status quo only), 3,063 conflicts of size 2 (i.e., two edits are in conflict, as well as with the status quo), and 140 conflicts of size 7 and more. We compare the average cross-entropy of the WoW(XLT) model with that of the random predictor and that of the WoW model. In Figure 5.5, we see that although the loss generally increases with conflict size for all three models, it increases less rapidly for the WoW(XLT) model than for the WoW model. This suggests that the explicit, latent, and text features enable the model to exploit the increasing complexity of data points to make more accurate predictions. We also see that for conflicts of size 4 and higher, the WoW model performs worse than the random predictor, but the WoW(XLT) model is able to outperform it.

5.4.7 Solving the Cold-Start Problem

We explore how to solve the cold-start problem by defining a second predictive problem: Given a dossier *i for which we have never seen an edit*, and given a conflict $\mathcal{C} = \{a, b, \ldots\}$, we want to predict which of the edits or the status quo would win. We order the dossiers by the date a committee receives a proposal, and we use the dossiers that contain the first 80% of the conflicts as a training set. We use the next 10% as the validation set, and we keep the last 10% aside as the test set. We ensure that no edits in the training set leak into the validation and test sets. This scenario is more realistic because we make

Table 5.5: Average cross-entropy of the baselines and our model on predicting new, unseen dossiers.

Type	Model	Avg. cross entropy
Baseline	Naive	0.947
	Random	**0.800**
	WoW	0.873
Ours	WoW(*Explicit*)	0.784
	WoW(*Text*)	0.839
	WoW(XT)	**0.759**

predictions about new dossiers that the model has never observed before.

We report, in Table 5.5, the results for WoW(*Explicit*), WoW(*Text*), and WoW(XT), together with the baselines. The latent features cannot be used for this task, as the dossier embeddings $\mathbf{y_i}$ are unavailable for new dossiers. The difficulty parameter d_i is set to the average difficulty learned in the training set. The random predictor, which learns the prior probability of the status quo winning for each conflict size, performs the best out of all the baselines and outperforms WoW(*Text*). Our approach outperforms the random predictor only when including explicit features. This suggests that the dossier features help us make more accurate predictions by learning parameter values for the type of dossier, its legal act, and its committee in charge. In this case, adding text features further boosts the performance.

The overall performance, however, is mixed: The improvement of WoW(XT) over the random predictor is rather small. One possible explanation is that the legislative process might be non-stationary. Hence, our model overfits the training set that is very different from the test set. The task is also unfair to our model, as in a real setting, predictions would be made only for the next dossier. In the current setting, we make predictions for all future dossiers. We keep further investigations of this aspect for future work.

5.5 Related Work

This work extends the dataset of Kristof et al. (2020) by including metadata features from the MEPs, the edits, and the dossiers, and text features from the edits and the title of the proposals. We augment their model by including these explicit features and text features into the WoW model. To strengthen the model, we also borrow from collaborative filtering techniques in the recommender systems literature. Similarly to matrix factorization techniques (Koren et al., 2009) that learn latent features for users and items to make recommendations, our model learns latent features for the MEPs and dossiers to predict edit outcomes. We show that these latent features improve the

predictive performance of our model by capturing bi-linear interactions between the MEPs and the dossiers.

Amendment analysis in the European Parliament has been studied by the political science community on datasets of small size (Baller, 2017; Kreppel, 1999, 2002; Tsebelis et al., 2001). Predicting edits on collaborative corpora of documents has been studied in the context of peer-production systems, such as Wikipedia (Adler & de Alfaro, 2007; Druck et al., 2008; Sarkar et al., 2019) and the Linux kernel (Jiang et al., 2013; Yardim et al., 2018). In this work, we combine the two by taking a peer-production viewpoint on the law-making process, and by proposing a model of the acceptance of the legislative edits. Our approach generalizes to any peer-production system in which the features of the users and items can be extracted and in which edits can be in conflict with one another.

We use the text of the edits and dossiers as features for classification. Text classification is a well-studied problem in natural-language processing. A simple baseline is to apply linear classifiers to term-frequency inverse document-frequency (TF-IDF) vectors (Joachims, 1998). However, these models do not capture the synonymy relation between words hence suffer from poor generalization. Models based on neural networks show better performance on this task (X. Zhang et al., 2015). However, they tend to require larger datasets, and the features they learn are harder to interpret. The fastText (Supervised) model (Joulin et al., 2017) bridges the gap between the two: It learns embeddings from linear models. We adapt this approach to our problem of edit classification, as edits are inhomogeneous pieces of text. Edit modeling has been studied using neural models(Guu et al., 2018; Yin et al., 2018) that suffer from the aforementioned issues of dataset size and interpretability. In the WoW models, we combine text features and non-text features to take into account the dynamics of the legislative process. Legal texts also have features and structures that set them apart from other domains. For example, the word “should” has a strong legal significance, whereas it is commonly removed as a stop word.

5.6 Summary

In this chapter, we extended our previous work on predicting legislative edits, where we considered influence parameters of the MEPs, controversy parameters of the dossiers, and the rapporteur advantage. We complemented our dataset with (a) additional explicit features of the edits, of the MEPs, and of the dossiers, (b) latent features of the MEPs and dossiers, and (c) text features of the edits and dossiers. Each of the three classes of additional features improve the performance significantly, and the best performance is achieved by combining all features. We interpreted the values of the learned parameters to gain insights into the legislative process. We provided interpretations of all explicit features to characterize the features that makes the success of an edit more likely. We showed that the latent features capture the representation of MEPs and dossiers in an ideological space. We analyzed the words and bigrams in different parts of an edit and

a dossier in terms of their influence on the acceptance probability. We also analyzed the performance of our model on subsets of the test set based on conflict size, and we showed that our best model can exploit the features of the data to make more accurate predictions on conflicts of higher size than other baselines. Finally, we described how to use our model for predicting edits made on new, unseen dossiers.

Ethical Considerations An anonymous reviewer expressed concerns regarding the use of machine learning for making decisions in law making, and whether our findings in Section 5.4 could help the perpetrators of adversarial attacks. We wish to clarify that we do not propose to rely on our models for making decisions, such as whether an edit should be accepted or not. Our goal is to understand the factors correlated with the acceptance of edits hence to gain insights into the law-making processes. These correlations do not imply a causal relationship that would benefit potential adversarial attackers.

Applications and Broader Impact We believe that approaches such as ours are helpful to political scientists, journalists, and transparency observers, and to the general public: First, it could be useful in validating theoretical hypotheses by using large-scale datasets and advanced computational methods. Second, it could help uncover lesser-known facts, such as controversial dossiers that slipped under the radar. Finally, the greater transparency that results from these insights can enhance trust in public institutions and strengthen democratic processes.

Future Work First, we currently use pre-trained word embeddings and embeddings trained on an ad-hoc binary classification task. We plan to explore how to learn text embeddings in an end-to-end manner by using the conflicting structure of the WoW model. Second, as shown in Section 5.4.7, our model has only a limited predictive power on edits made on future dossiers. We plan to further explore how to exploit the temporality of the data and how to develop a dynamical model able to take into account the non-stationarity of the law-making process. Finally, we plan to explore more complex models of textual edits, such as considering pairs of words that are inserted and deleted and longer-range word order.

6 Conclusion

In this thesis, we have explored different parts of the socio-political system in representative democracies and have answered research questions stemming from the problems that affect their proper functioning. Given the recent availability of vast amounts of digital data in this domain, we took a computational and data-driven approach and built interpretable models of social phenomena as a means to answer these questions. As language is a ubiquitous modality of social data, we based our models on methods from NLP, and to capture human preferences and to estimate subjective quantities, we then incorporate concepts from discrete-choice theory.

In Chapter 2, we built text-based models to score subjective bias in web documents such as Wikipedia articles and news media. By framing the problem as a pairwise comparison of bias, we were able to benefit from larger and better-quality training data. This in turn enables us to build simple and interpretable models that achieve a good level of accuracy comparable to that achieved by deep neural networks and humans. We were able to discover the words indicative of bias by using the learned model parameters. Although it was trained for pairwise comparisons, our model computes real-valued bias scores for individual documents. We use these scores for several applications, including tracking the evolution of bias in Wikipedia articles throughout their history, comparing bias in media outlets, and scoring the bias in political speeches, law amendments, and tweets. In each case, we have shown that the scores correspond to expected patterns of bias, and we have provided interesting new insights into the manifestation of bias in social settings.

In Chapter 3, we have studied effective communication strategies for maximizing user engagement in social media campaigns, taking the example of campaigns about climate change on Twitter. By comparing the engagement of pairs of tweets that are from the same author and made around the same time, we were able to avoid confounding factors and to learn interpretable models for predicting engagement, based on the tweet's topic and metadata features. Based on features thus discovered to be correlated with high engagement, we have made recommendations to optimize communication about climate

change.

In Chapter 4, we have turned our focus to the political side of the system and have explored methods to shed light on the influence of lobbies on the law-making process in the EP. We curated a rich dataset of lobbies' position papers by crawling their websites and matched them to the speeches and law amendments made by MEPs. We validated the MEP-Lobby links obtained by comparing them with a curated dataset of retweet links between the two entities and with the publicly disclosed meetings of MEPs. We also made an aggregate analysis of the links and observed that the patterns match as expected, based on the ideology of MEPs and on the area of work of the lobbies.

Finally, in Chapter 5 we have focused on the law-making process *within* the EP. To understand the factors correlated with the success of law amendments proposed by MEPs, we built interpretable models that predict the acceptance of amendments within parliamentary committees. Our models incorporated explicit features of the amendments, MEPs, and laws, text features of amendments and laws, and latent representations of MEPs and laws. We have discovered interesting factors associated with amendment acceptance: These factors include the status of the proposing MEP being the committee reporter, the presence of the optional justification for the amendment, the text suggesting a watering down of the law, such as the addition of recommendations ('should' rather than 'must'), and the narrowing of the scope of the law such as the addition of bigram 'where applicable' and removal of bigram 'any other'.

In this thesis, we have demonstrated that interpretable text-based models can be constructed to understand the different phenomena in socio-political systems of representative democracies. This can, in turn, help solve some of the problems affecting the effectiveness of these systems. Improving the transparency of law-making processes enhances the accountability of institutions, such as parliaments, and helps to increase citizens' trust in them; this trust is the bedrock on which democracies function. Tools for measuring subjective bias on the web and media can help citizens make better-informed decisions, and effective communication strategies can help motivate them to take action for social causes such as climate change.

Ethical Considerations Working with social data necessarily involves ethical considerations, as it is human-generated. We have mentioned several relevant points at the end of each chapter. Here, we briefly summarize these points and make some general observations. We used only publicly available data in this thesis. We did not use any personal data, other than tweets and MEP data. In the case of tweets, we followed Twitter's Terms of Service and obtained data through their official API. As MEPs are public officials, the use and reproduction of their data on official websites, including speeches and amendments, is authorized by the EP.

Wikipedia data is released under CC BY-SA and GFDL licenses and the analysis of it does not require informed consent. Lobby documents from different websites were crawled in parallel by using a cluster, which allowed us to use a reasonable interval between successive requests to each website hence did not affect their regular operation. We do not release copies of original lobby documents in order to respect copyrights; we release only the GPT-generated summaries (as allowed by OpenAI's Terms of Service), URLs to the original documents, and archived versions on Internet Archive where possible to mitigate link rot. We ensure that the summaries of position papers that we release do not contain any personal data.

Care is needed while drawing conclusions based on the output of machine-learning models, as they can occasionally make surprising mistakes. For instance, our bias scoring model in Chapter 2 could assign a high bias score to a text that is actually fairly neutral. The lobby-MEP association model in Chapter 4 could assign a high association score to a pair that has little or no association in reality. However, the *interpretability* of our models mitigates the effects of this issue to some extent. The user could examine the words that the model considers to be biased or the documents that the model considers to be similar, and re-assess the model's decision.

Future Work There are several directions to take to develop the work presented in this thesis further. We briefly outline them in the following paragraphs.

Our *datasets* can be expanded. In the current work, we focused only on English texts, but it is straightforward to extend our analyses in Chapters 2, 3, and 4, and to include text in other languages. The temporal range of climate-related tweets we studied in Chapter 3 could be extended and the difference in communication strategies across time periods could be studied. It could also be interesting to study the strategies employed by climate-change deniers so as to devise effective countermeasures. The recall of the position paper classification step in Chapter 4 could be improved to provide better coverage of the lobby positions. Documents from the lobby websites, which have been archived by the Internet Archive, could also be included to improve coverage of the positions taken on issues in the past.

Our *methods* can be improved. Although we restricted our bias scoring models in Chapter 2 and our edit success prediction models in Chapter 5 to using fixed pre-trained word embeddings to keep the computational cost manageable, we could fine-tune the embeddings together with the models to potentially obtain better accuracies. We could also train generative models to automatically reduce the bias in text using the minimization of scores computed by our bias scoring models as the training objective. We could extend our engagement prediction model in Chapter 3 to consider audience characteristics and trending topics. The aspects of climate change that increase user engagement could depend on the interests of the users in question, for example, users

interested in wildlife could be more likely to respond to themes such as habitat conservation rather than energy efficiency; and could depend on the trending topics at the time the tweet is made, for example, around the time of a natural disaster, tweets about it and related topics could encourage higher engagement. We could link lobbies to the amendments they support by using the methods in Chapter 4 and compute skill parameters and latent representations of lobbies by using the predictive models of amendment success in Chapter 5.

Useful *applications* could be built on top of our models. The Predikon project (Kristof, 2021b) for predicting Swiss votes, which is based on the work in Kristof (2021a) is a source of inspiration in this regard. A straightforward application, based on the model in Chapter 2, could take in a piece of text and display its bias score, together with its percentile score compared to the bias in Wikipedia articles. The tool could also highlight words in the text that make the highest contributions to the bias of the document. This could be useful for Wikipedia editors to monitor and to flag biased content, for authors to draft unbiased text, and for users to understand the level of bias in the news that they are reading. A similar tool, based on the engagement prediction model in Chapter 3, could help authors draft effective messages on social media. The data and models in Chapters 4 and 5 could be used to build powerful tools for enhancing the transparency of EU institutions. For instance, links between MEPs and lobbies could be displayed through interactive graph visualizations, along with the specific speeches, amendments, and lobby documents that each link is based on.

A Social Media Campaigns

To get a more detailed understanding of the topics, we give here a sample of 10 tweets chosen uniformly at random from the top 500 tweets that have the highest probability for each of the topics in Figure 3.1. Author names are hidden and special characters are removed. The Title and Description of linked pages in URLs have been expanded where available. Note that some of the tweets may contain offensive language.

President

- Biden plans to fight climate change in a way no US president has done before httpstcoGgemljiccs
- GinaMcCarthy White House National Climate Adviser discusses Pres Bidens infrastructure plan httpstcozy24j9Cxuk httpstcoLOWQgcDdb9
- President Joe Biden has already taken more executive action on climate change and the environment in his first week in office than any president before him httpstcoFMpGOz6GSt
- Biden will issue executive orders rolling back Trumps climate policies The Washington Post httpstcon2oCIYmu6X
- Former VP Gore Begged Biden Not to Compromise Climate Change in Bipartisan Infrastructure Deal httpstcoBTd6n17z4g httpstcoFSmMV3TCZh via Newsmax Title Newsmaxcom Breaking news from around the globe Desciption Newsmaxcom reports todays news headlines live news stream news videos from Americans and global readers seeking the latest in current events politics US world news health finance and more
- Joe Biden announces he has cancer during speech about global warming httpstcoDiAUXSb7S8 Title Joe Biden announces he has cancer during speech about global

warming Desciption JOE BIDEN announced he has cancer during a speech about global warming

- httpstcouqTL9dePmJ In his first address to Congress President Joe Biden touched on immigration police reform and climate change httpstcoT7Nt3ksSRc
- Trump dismisses climate change calls on Biden to fire joint chiefs httpstcoFOaxDK6X3O httpstcoJN8fNQEFpg
- WATCH LIVE Biden signs climate change executive orders httpstcoREgDK4ntJq
- Biden will issue executive orders rolling back Trumps climate policies The Washington Post httpstcoSUC6GLxQAi

Clean Energy

- Electricity production from fossil fuels nuclear and renewables Switzerland httpstcoApAzM7YM6c
- Three major categories of energy for electricity generation R fossil fuels coal natural gas amp petroleum nuclear energy amp renewable energy sources Most electricity is generated with steam turbines using fossil fuels nuclear biomass geothermal and solar thermal energy
- In 2021 40 of the Electricity Produced in the United States Was Derived from NonFossil Fuel Sources CleanTechnica httpstcoiw0Y1sTIO0 Title In 2021 40 of the Electricity Produced in the United States Was Derived from NonFossil Fuel Sources Desciption In 2021 40 of the Electricity Produced in the United States Was Derived from NonFossil Fuel Sources reliance on fossil fuels
- Nuclear Power The Reliable Energy Source Nuclear power energys ability to produce electricity with lower carbon emissions than fossil fuels has been driving the markets development Get More info httpstcoCAPZIDibUV nuclear nuclearpower energy usak France russia httpstcokiwQvSyRue Title Nuclear Power Market Global Size Share Industry Trends 2022 2028 IMR Desciption Global Nuclear Power Market was valued at USD 7141 Billion in 2021 and is projected to reach USD 16347 Billion by 2028 growing at a CAGR of 1256 from 2022 to 2028
- Portland General Electric an Oregon utility is poised to build the nations first largescale wind solar and battery facility to accelerate the transition from fossil fuels to 247 cleanenergy httpstcoxPxHCIX1Zb Title Oregon utility powers up nations first largescale wind solar and battery facility Desciption Portland General Electric has built a firstofitskind facility that will use an innovative battery technology supporters are calling a game changer for Oregons renewable energy transition

- Upscaled renewable energy capacity will improve the resilience of Australias main electricity grid but outages from fossil fuel plants the main threat to supplies httpstcowtT6GtDlYj httpstcokcPW92WMOA Title Fossil fuel plant outages pose main threat to summer power supply as renewables bolster grid Desciption Australian Energy Market Operator says risks of insufficient supply during summer peak load periods remain despite La Nia bringing cooler temperatures

- The Midwest is preparing for a fossilfuelfree future new transmission lines will stabilize the grid as it shifts to renewables It will facilitate retirement of more than 50 gigawatts of fossil fuelpowered electricity primarily from coal plants httpstcoU7j8yv0VXt Title Biggest story of the year for renewables Desciption The Midcontinent Independent System Operator approved a 103 billion investment in 18 new transmission lines this week Never before have so many power lines been approved all at once

- More renewables than fossil fuelbased electricity in the grid pv Europe httpstcoYU3MlDk6eI

- New post Renewables Overtake FossilFuel Power Plants In EU Electricity Generation httpstcoXviCd5qHxS

- The US Energy Information Administration has forecast in its January ShortTerm Energy Outlook that rising electricity generation from clean energy such as solar and wind will reduce generation from fossil fuelfired power plants over the next 2 years httpstcoCDIpgohIl5 httpstcoykeMKGuPks Title NonRenewable Energy

Drought-resistant plants

- 7 droughttolerant houseplants to add some greenery to your home httpstcoQN9B71d6an httpstcozK8XLtUOQ3

- Drought tolerant soybean variety httpstcolKfheA21CB

- Tomatoes and Other Crops Wither in California Drought httpstconZRUevvzhH Title Tomatoes and Other Crops Wither in California Drought Desciption The Western drought has come for pasta sauce and ketchup Processing tomatoes used in innumerable grocery store staples are suffering from years of subpar rainfall and snowpack in California The

- Rediscovering ancient crop varieties httpstcojnWo4iLBLZ

- Diversity Among Synthetic BackDerived Wheat Triticum Aestivum L Lines for Drought Tolerance preprints httpstcohp8ctHMLhL

- Chile drought fruit crops under threat httpstcoq0YsYcxSi4 httpstcowkbenXmfQX

- Scientists developing drought and wet tolerant soybean varieties httpstco2I6IM9fZ29 Title Scientists developing drought and wet tolerant soybean varieties Talk Business Politics Desciption Drought has been a major problem for Arkansas farmers this year and if climate prediction models are correct it will be a major problem in the coming years and decades
- Grafting is a technique used to grow different types of fruits and vegetables For tomatoes a desirable fruitbearing scion the shoot is placed atop a rootstock the roots that contains some other desirable characteristic httpstcoZyESGFyMdz camerawithflash Steven Bristow httpstcoYxcEmu0nwT Title How do rootstocks help tomato growers under heat and drought Desciption No matter the location plants can experience stress This can be in your home garden the local community garden or on a farm miles away Stresses include heat or drought and they limit a plant
- Echium vulgare a very drought resistant plant httpstcoiuSgLiC8Be
- Top seller in Jan snowflake Cold Start snowflake Germinates from 5 degrees High wear tolerance drought tolerant fine leaved httpstcoTWDdNVhVsd

Africa

- The Horn of Africa is facing drought and food shortages httpstcorvez2vHA0q Title The Horn of Africa is facing drought and food shortages Desciption Ayesha Rascoe speaks to Samantha Power administrator of the United States Agency for International Development about food shortages and drought in the Horn of Africa
- Somali presidents envoy for drought Abdirahman Abdishakur and his delegates visited displaced people inthe IDP camps in the outskirts of baardheere town Gedo region today AAbdishakur HassanSMohamud httpstco4zZk0UjI93
- Horn of Africa faces brutal drought and food crisis httpstcoI0gvTYzPtj httpstcozQXsR8naVh Title Horn of Africa faces brutal drought and food crisis Citypress Desciption The latest outlook confirms the fears of aid agencies which have been warning for months about the worsening consequences of the drought in Ethiopia Somalia and parts of Kenya
- Buhari to attend conference on desertification drought degradation in Abidjan httpstcowQvWB0Fk3c News Abidjan buhari BBnaija Messi Wizkid Davido
- Drought is on going in Angola Congo DRC Namibia httpstcoAwfWGUzoBF
- Thousands of Ethiopian refugees who have fled Ethiopias Tigray region are now struggling against extreme weather in eastern Sudan cloudwithlightningandrain httpstcomiCEESE0pt httpstcoXATNkiQJZ0

- DROUGHT AND FAMINE BITE HARD IN KAKUMA REFUGEE CAMP UHAI EASHRI Refugee Coalition of East Africa UNHCR theUNRefugeeAgency Bisoea TheTaalaFoundation TheActionFoundation TeamNosleepfoundation KenyaRefugeeLaw ProjectTransAdvocacy Initiative ORAM httpstcoZSbsztfJtW
- 35 MILLION Kenyans are facing food shortage due to ongoing drought IGAD executive director Dr Workneh Gebeyehu says httpstcoouwZge9a8e
- Not Kaizer Chiefs fans blaming Komphela for their trophy drought Kubi facewithtearsofjoy httpstcoeERuQ9cJl2
- World renowned Oromian athlete of the day she donated 4 mil birr to drought victims in Bale Oromia May Waaqa bless you and Oromia httpstcoHnb59jJPbx

Planet

- earth planet of life is destroyed by global warming pensiveface
- Life may have thrived on early Mars until it drove climate change that caused its demise space httpstcopbIKibxpaC Title Life may have thrived on early Mars until it drove climate change that caused its demise Desciption If there ever was life on Marsand thats a huge ifconditions during the planets infancy most likely would have supported it according to a study led by University of Arizona researchers
- Ancient dinosaurs went extinct 65million years ago due to planetary catastrophe and climate change from highspeed collision with enormous astroid causing widespread creator explosion
- The titanoboas existence was made possible by a warming planet httpstcoNRHrCl0HlE
- Astronomers discover potential water world exoplanet nearby Earth httpstcoE8Ojx1w1vr via Yahoo Mankind is for the foreseeable future confined to Earth not committed to keeping it livable for themselves still use fossil fuels have yet to achieve lightspeed capability Title Astronomers discover potential water world exoplanet nearby Earth Desciption NASA says the exoplanet located just 100 light years away could be a water world
- How superhot rocks miles under the earths surface could provide limitless clean energy httpstcotgzYqkjUqK Title How superhot rocks miles under the earths surface could provide limitless clean energy Desciption Superhot rock geothermal energy can be generated from dry rock thats at least 752 degrees Fahrenheit which lies at depths between two and 12 miles
- Cryptocurrency is the intermediary between a fiat controlled divided fossil fuel operating society and a decentralized unified electromagnetically propelled society

As long as the Illuminati control the economy and subsequently everything else Earth humans cannot be free

- Jeff Bezos and Elon Musk rather than spending billion of dollars traveling to the space use the money instead to save the planet earth from destruction by global warmingclimate change
- The nations of the Global North have effectively colonised the atmospheric commons Theyve enriched themselves as a result but with devastating consequences for the rest of the world and for all of life on Earth httpstcoDaFOqaPzWk Title Why climate change is inherently racist Desciption Climate change divides along racial lines Could tackling it help address longstanding injustices
- In 60 million years we will be the dinosaurs who destroyed ourselves with nukes and climate change our fossils studied by some future intelligent species

Fossil Fuels

- Buy electric cars to eliminate fossil fuels Electricity is made through fossil fuels such as oilnatural gas and coal facewithtearsofjoyfacewithtearsofjoyfacewithtearsofjoy
- There is no such thing as a zero emission vehicle because the electricity to fuel such vehicles and battery require the burning of fossil fuels EVs will NEVER replace their fossil fueled counterparts
- Fossil Fuels vs Renewables ALL Forms of Energy are Intermittent httpstcoAeGIYC9It3
- Without fossil fuels renewable energy would not exist
- Could carbon fibres soon become free of fossil fuels httpstcoUW13DTfA8r
- Fossil Fuels Recycling Wealth
- Electric vehicles increase demand for fossil fuels
- Reliance on fossil fuels must be phased out GovKathyHochul
- Electric cars run on fossil fuels nuclear power is the answer
- Fossil Fuels vs Renewables ALL Forms of Energy are Intermittent httpstcooeru25EXAH

Politics

- Six or more core liberal seats ripped away thanks to scomo running a republican one nation lite platform and ignoring climate change Great result

- Stop Trump Republicans from seizing control of California by voting NO on the recall Front runner is lunatic republican Larry Elder antivax conspiracy spreader and climate change denier VoteNoOnTheRecall

- Dem says Manchin blocking energy tax provisions in big bill httpstcoAZDKY6FnJr Title Dem says Manchin blocking energy tax provisions in big bill Desciption Sen Joe Manchin has told Senate Majority Leader Chuck Schumer that he will oppose a economic measure if it includes climate or energy provisions or boosts taxes on the rich or corporations

- In Florida Dems plan to go after Marco Rubio for voting no on the massive reconciliation bill that deals with climate change health care and taxes httpstcoyqyaIxsrTn Title Dems hope to punish Rubio over Senate vote Desciption Big vote cometh The Senate approved its massive reconciliation bill that deals with climate change health care and taxes on Sunday and as expected Sens Marco Rubio and Rick Scott voted no On the home front Theres the broader political question of whether the bill will help Democrats and President Joe Biden in November But how will it play in Florida Both sides tried hard over the weekend to make the votes on the legislation which heads to the House next as well as individual amendments

- Dem says Manchin blocking energy tax provisions in big bill httpstco59NB1ClSDI Title Dem says Manchin blocking energy tax provisions in big bill Desciption Dem says Manchin blocking energy tax provisions in big bill

- Lib Dems call on PM to act local on climate change httpstcoYcbNfEYziW httpstcoov5kH8ytbm

- Liberal MPs consider crossing the floor to support climate change legislation httpstco8SHG3qcCle Title Coalition says it will oppose governments plan to legislate emissions reduction target Desciption The Coalition has confirmed it will oppose Labors plan to enshrine in legislation a 43 per cent emissions reduction target by 2030 after holding a party room meeting

- BREAKING NEWS Key Biden Nominee Open to Raising Taxes on Middle Class to fight for Climate Change httpstcoj6N4otIl7g UnitedStates

- Democrats Want Biden To Go Beast Mode amp Fight ClimateChange Via Executive Action httpstcoGOX65U2E1q verbalese Shi4Tech PeterSBecks1 freyyjaa88 BettyFellows ableraces danigirl1207 RaymondNorman AKimCampbell versusplus thumperftw bugs4US MichaelChrisLA httpstco0NE1MlilYt Title Democrats Want Biden To Go Beast Mode And Fight Climate Change Via Executive Action Desciption Time is of the essence if the US wants to avoid a global climate catastrophe Democratic senators warned after hopes for climate legislation faded once again

- Slimmed Down Energy Tax and Social Spending Package Targeted for Vote Before August httpstco1Ytdz5Bhn1

Global Warming

- Why is it still hot in November Global warming is terrifying
- If this is global warming then I love it warm weather then warm rain
- The weather today is super hot is it because of the global warming or me being effortlessly hot
- Its a hot humid August day in Texas Must be climate change
- How to keep your home cool in the heatwave in preparation for the rising temperatures httpstcozjVrBcyJqr httpstcoVdZ2l6oHcc
- This country is not prepared for extreme weather of any kind hot cold stormy womanshrugging
- I remain cool until global warming make me hot firefirefire httpstcolWxb71NFj8
- Climate changes freakishly hot summer is leading to a new type of insurance coveragefor heat stroke httpstcoizdt8kKaob Title Climate changes freakishly hot summer is leading to a new type of insurance coveragefor heat stroke iTech News Desciption Business is booming for heatstroke insurance providers in Japan One company saw a 1600 increase in sales of daily coverage during the final week of June Read More Source Business Fortune
- If global warming isnt real why am I sweating in January facewithmonocle it shouldnt be this warm in January fam
- This has to be one of the hottest driest summers ever Global warming is real

Geopolitics

- Saudi Arabia Japan and Australia are among countries asking the UN to play down the need to move rapidly away from fossil fuels httpstcoNW68HQbBaX Title COP26 Document leak reveals nations lobbying to change key climate report Desciption Countries are asking the UN to play down the need to move rapidly away from fossil fuels
- Water woes caused by climate change could compel Iran to seek a deal on its nuclear program and join others in the Middle East in water cooperation httpstcorLxcXE6GLx

- Reuters US President Joe Biden on Friday called on China and other major economies to redouble their efforts to combat climate change and improve energy security warning that Russias invasion of Ukraine had sharpened the need for urgent action

- China stands firmly with Pacific island countries on national sovereignty and security climate change maritime rights and interests said Chinese State Councilor and Foreign Minister Wang Yi httpstcoD6qA6KH7Tx

- AustraliaChina IRENA and the MEE of the Peoples Republic of China have extended existing cooperation on energy transition httpstcoyOLIwWJdjm

- Alaska talks reveal the tense relationship between US and China on climate change httpstcooJ5p2iogsQ

- Obama faults Russia and China for dangerous lack of urgency on climate change httpstco079770LtDN

- NATO SecretaryGeneral Jens Stoltenberg says climate change and Russian aggression mean defence of the Arctic is key httpstcoVa3CtBzwZW httpstcoqLTuWlnoTA Title Stoltenberg Arctic key in defence against Russia Desciption NATO Secretary-General Jens Stoltenberg says climate change and Russian aggression mean defence of the Arctic is key

- Washington sees Germany as a critical ally on everything from confronting China and negotiating with Iran to climate change Solving their NS2 problem would allow them to focus on their shared agenda FT httpstcoeACl4OW30v

- Australia is the weakest link in international sanctions on Russia crikeynews httpstcoCKzPyUZw4d

Anger

- This is infuriating enragedface Screwing the planet to line their pockets and easily selling it all to a gullible constituency httpstco6JtPick1Qg Title Republican states are trying to force people to keep using fossil fuels Desciption Most people recognize that we need to wean our way off fossil fuels and embrace clean energy if we have any chance of avoiding the worst effects of climate change Unfortunately some of the people who dont realize this reality happen to be the

- Republicans have spent the last four years calling climate change activists the radical left meanwhile theyve literally created a coup full of conspiracy theorists and racists with guns who are trying to breech the Capitol as we speak to tear down democracyTHAT is radical

- These green fanatics are nothing more than the footsoldiers of the elites who invented the man made climate change hoax to control people and not to control the climate jeremyvine httpstcoY7oocc4XtX
- AGW Climate Change or whatever they want to call it is nothing more than a way to bilk gullible people into forking over Trillions of hard working peoples tax dollars to a multinational MONEY LAUNDERING SCHEME
- Fucking despise the Tories ensuring we are world beating at fucking the planet up httpstcoaZuEnnXvks Title Fury as government overrules council to approve absurd Surrey gas drilling Desciption MP Jeremy Hunt blasts decision that goes against local opinion and government commitment to devolution of powers and causes enormous anger
- US castigates China over climate change efforts httpstcowghTwnLouT via BBCNews Beware you cannot trust the inhuman Communist Government of China No respect for these so called human enslavers
- Liberals are corrupt stupid bastards httpstcoBL6c8sB06Y
- These Climate Change freaks make me sick Biden and the Democrats are among these filthy trashruining the US economy because they believe they are righteous In reality they are little more than misguided spoiled brat lazy useless punks stealing from the workers
- The gall is staggering They are criminals who are willfully destroying our world amp should be treated as such httpstcoTYQsNgDPG2 cop26 JoeBiden JohnKerry BorisJohnson ScottMorrisonMP EmmanuelMacron EPAMichaelRegan epa JustinTrudeau DemWarRoom HouseDemocrats ap
- The weatherchannel just lost my viewership Theyve become the purveyors of the liberal climate change is based on racism ideology Shame on you for politicizing amp dehumanizing people of color

Geology

- Supervolcano Eruption Depleted Earths Ozone Layer 74000 Years Ago httpstcoCqpoogQAzd
- Volcanic surge narrowed seas during ancient global warming event Climate GlobalWarming Volcanos httpstcod7rI4KeDpx via physorgcom
- Past AMOC collapses were caused by warming of ocean subsurface amp reduction of surface salinity due to icebergs separating from glaciers into the sea Uh oh grimacingface BlackBearNews1 Blueoceanarctic newday2020sc WaveFoundation Kameshwarikate httpstcoXrw4VRhliC Title Global warming could collapse the

Atlantic circulation system Desciption Global warming could collapse the Atlantic circulation system Earthcom

- Tipping points in Earths system triggered rapid climate change 55 million years ago research shows PhysOrg httpstcoBH5XvDQ9Jp httpstcoZvJlp6Aq58
- Shifting Signatures of Climate Change Reshuffle Northern Species httpstcoU6TdPD6Aw4 Title Shifting Signatures of Climate Change Reshuffle Northern Species Current Science Daily Desciption Analysis of longterm monitoring data for almost 1500 species in Finland shows that four decades of climate change has led species to shift between the better and worse parts of their climatic niches and that these impacts were most pronounced at higher latitudes
- Antarctic Circumpolar Current flows more rapidly in warm phases In future the intensity of the Antarctic Circumpolar Current could increase accelerating climate change httpstcoOzkbB5r1y4
- Arctic Animals Movement Patterns Shifting Due to Climate Change httpstcoiQ39qXuTxW httpstcozibpRCLelS
- A climate warming event 56 million years ago resulted from the release of greenhouse gases most likely from a volcanic eruption httpstcou4PqcuGumN
- Tipping points in Earths system triggered rapid climate change 55 million years ago research shows httpstcoEtqXhcvVC1 httpstcoNHdC7alKOz Title Physorg News and Articles on Science and Technology Desciption Daily science news on research developments technological breakthroughs and the latest scientific innovations
- 15 C Cap Could halve Sea Level Rise From Melting Ice Study httpstcobC7I5dTKJY

Mixed

- Our atmosphere is shrinking httpstcoZN4hmHiIGb httpstcoTzQekR8iEe
- Baetokkis survived the renebaebae droughtI hope shell become more active soon
- How the Moon Wobble Affects Rising Tides Ecology astronomy via httpstcoB2dOvBHk3D httpstcoxicR96Q3h8 Title Twinybots Empower your business
- our fandom is probably the most wellfed fandom on all of stan twitter yet ariana doesnt release music for two weeks and yall start acting like were in a drought loudlycryingfaceloudlycryingfaceloudlycryingfaceloudlycryingface
- i just realized now that josh is done releasing songswe might go into a content drought
- hes currently saving stays from instagram drought LMAOOOO httpstcoYD3GwIU1QR

- Australia faces constellation of diplomatic pressures over climate httpstcoEPX83YQ94O
- Latest climate summit draft waters down fossil fuel language httpstconN4OpAJBoY ClimateEmergency COP26 JoeBiden XiJinping ClimateCatastrophe ClimateActionNow EnergyTransition GretaThunberg
- did yall know that rsl is ending the acting drought for Me exclamationquestionmark
- New study uncovers hidden behaviour of the Arctic Oceans currents that could alter future climate change predictions National Oceanography Centre httpstcoZMDaPmOkwi

Low water

- The Hoover Dam reservoir is at an alltime low httpstcofaboCIXwap
- Lake Mead falls to an unprecedented low exposing one of the reservoirs original water intake valves httpstcos8vRKL45o8 Title Lake Mead plummets to unprecedented low exposing original 1971 water intake valve Desciption Lake Meads plummeting water level has exposed one of the reservoirs original water intake valves for the first time officials say
- In dry California salty water creeps into key waterways from AP httpstcor2Mukxg8eA Title In dry California salty water creeps into key waterways Desciption RIO VISTA Calif AP Charlie Hamilton hasnt irrigated his vineyards with water from the Sacramento River since early May even though it flows just yards from his crop Nearby to the south the industrial Bay Area city of Antioch has supplied its people with water from the San Joaquin River for just 32 days this year compared to roughly 128 days by this time in a wet year
- Houseboats removed from LakeOroville which stands at 38 percent of capacity CAwater CAdrought California water drought SaveOurWater httpstcoTQKC4K4TxZ
- The Hoover Dam reservoir is at an alltime low httpstcoxdZxAggtWa httpstcow4d2ZHUcaO
- Greenhouse gas dynamics in an urbanized river system influence of water quality and land use httpstcoYlx25xKnDO
- Low water levels due to drought reveal sunken car in Pineview Reservoir Salt Lake Tribune httpstcoyb8gpqK6U4
- Critically low water levels at Lake Shasta Californias largest reservoir httpstcoqMHkQGxLRJ Title Critically low water levels at Lake Shasta Californias largest reservoir Desciption Lake Shasta the largest reservoir in California is experiencing the lowest levels of rainfall since Shasta Dam was constructed in the 1940s Water

allotments to Central Valley fammers and urban water districts in the Bay Area have been severely cut back as a result

- The Hoover Dam reservoir is at an alltime low httpstcosINziskFrF
- Measuring the Bathtub Ring Calculating Reservoir Surface Area Changes in the Colorado River Basin planet SatelliteData Drought Reservoirs httpstco7BNqhrF7cT

Conference

- CONFERENCE 57th session of the Intergovernmental Panel on Climate Change IPCC 57 2730 Sep 2022 Geneva Geneve Switzerland httpstcozgFLPM6hMx Title Event 57th session of the IPCC IPCC 57 SDG Knowledge Hub IISD Desciption Tracking the Implementation of the 2030 Agenda
- POTUS and world leaders will discuss climate change on the February 19 virtual G7 hosted by BorisJohnson per PressSec
- New Post OPEC hosts coordination meeting on climate change httpstcoX5sHOjDaeQ OPEC hosts coordination meeting on climate change
- The UK hosts the 26th UN Climate Change Conference of the Parties COP26 Scotland Glasgow now httpstco0CaEsS9YRr
- COP26 Presidency Meeting with UNFCCC Observer Focal Points has held at the ongoing United Nations Climate Change Conference and censoj participated httpstco1dGvSO9uni
- The intergovernmental panel on climate change opens second draft on group III sixth assessment report httpstco6c2IK74YZZ
- The United Nations Framework Convention on Climate Change UNFCCC will hold its 26th Conference of Parties COP26 in November 2021 It will be hosted by the UKinUganda and will take place in Glasgow httpstco2dta51AAu2
- AIA will send representatives to the United Nations Climate Change Conference COP27 again this year httpstcogtFBKDOerM Title How AIA is helping to combat the climate crisis at COP27 Desciption For the second year in a row AIA is sending representatives to the United Nations Climate Change Conference COP27 an annual Conference that brings together government officials and nongovernmental organizations to collaborate on ways to combat the climate crisis
- New Post OPEC hosts coordination meeting on climate change httpstcobo71mMQ1Us OPEC hosts coordination meeting on climate change

- ICYMI In this IABC22 plenary session a panel comprised of climate change communicators will discuss the evolving approach to change the conversation Join the discussion at the World Conference 2629 June httpstcoCJVhU1vNEu httpstcoLiunP7Bj2Y Title Keynote and Plenary Sessions

Research

- A major finding using longterm consistent satellite climate data from the esa Climate Change Initiative httpstcopKDCSaoaXO

- A Global LISOTD Climatology of Lightning Flash Extent Density httpstcoRI3kBcdJl4

- BOE climate stress test results httpstcosM3xpvdTGo Title Bank of England publishes results of the 2021 Biennial Exploratory Scenario Financial risks from climate change Desciption The Bank of England Bank has today published the results of the Climate Biennial Exploratory Scenario CBES which explores the financial risks posed by climate change for the largest banks and insurers operating in the UK

- The analysis is the latest in a series of studies that show the influence of climate change on extreme weather httpstcoTRnNGcZO1e Title Did Warming Play a Role in Deadly South African Floods Yes a Study Says Desciption Climate change sharply increased the chances of catastrophic rains in the countrys east a team of researchers has found

- This study has improved our understanding of combined droughtheat wave events by considering a much finer temporal resolution than previous studies allowing a more refined consideration of risk forecasting and hazard preparation for such events httpstcosKZ4X2Pra5 Title Simultaneous Drought and Heat Wave Events Are Becoming More Common Eos Desciption As the world heats up the number and duration of combined stress events are increasing causing harmful environmental and human impacts

- OHC is an important indicator of the global climate change Here we conducted a comparative study of OHC among different data sets including observationbased ones Argoonly and Argoother observations ocean reanalyzes and freerunning model httpstcoCwyUcPFo09 Title A Comparative Study of the ArgoEra Ocean Heat Content Among Four Different Types of Data Sets Desciption Global and basinswide ocean heat content OHC trends were largely similar among the observationbased data sets Ocean reanalyzes RAs well captured the largescale warming and cooling patterns

- SUNGHOON DROUGHT rollingonthefloorlaughingrollingonthefloorlaughing

- According to new research from Proceedings of the National Academy of Sciences PNAS regarding climate change and pollen seasons there has been a significant increase of pollen in the air in recent years httpstcoQeBsZHcBoy

- A study found that 218 infectious diseases had been exasperated by climate change backhandindexpointingdown httpstcozvvrNVKUdf Title Climate change may be fueling infectious disease outbreaks Luke ONeill says Desciption A study found that 218 infectious diseases had been exasperated by climate change

- Video summary of the Copernicus Climate Change Services report httpstcob5XcjzKFqI

Youth

- Climate Change AnimationInteractive Poster Competition Open to all year groups BIG PRIZES Click link for details httpstcofHgGijVpdk httpstcoEd1mgfiCzM

- Our course Climate Change Resources for K12 Teachers and Students is for K12 teachers and students looking for great climaterelated talks videos games websites and more resources they can use in advancing their understanding of climate change httpstcoL7eqHUOLQI Title Climate Change Resources for K12 Teachers and Students Desciption Topical Content and Hundreds of Resources for Exploring and Learning About Climate Change

- We are creating 450 jobs for Youth The ScienceHorizons Youth Internship Program equips recent grads with great work experience in the CleanTech sector Apply today httpstcouds9GnYSBb httpstcorL0dPtvC7F Title Science Horizons Youth Internship Program Canadaca Desciption The Science Horizons Youth Internship Program provides wage subsidies to eligible employers across Canada to hire recent university college and polytechnic graduates for internships in the environmental science technology engineering and mathematics sectors Description objectives and contact information Application process for employers to receive funding Internship opportunities application process for postsecondary graduates Interactive map of past internships across Canada You will not

- Climate Change Writing Competition open to all UG and PGT students A perfect opportunity for our MA cohort to demonstrate their new WeAreALBERT knowledge httpstcobNzThDwL3n

- Join SPIDTheatre for the final month of exciting drama workshops as part of the Far Far Away online project in collaboration with sciencemuseum Available to 813 year olds who want to get involved with a performative project about climate change and the environment performingarts globeshowingEuropeAfrica httpstcozeNG2so53h

- DYK ScienceHorizons Youth Internship Program helps young Canadian graduates gain valuable work experience in STEM This will support their success in the clean energy job market Learn more httpstcoiZVO4eRYcL httpstcopSOECkpEw6 Title Science Horizons Youth Internship Program Canadaca Desciption The Science Horizons Youth Internship Program provides wage subsidies to eligible employers across Canada to hire recent university college and polytechnic graduates for internships in the environmental science technology engineering and mathematics sectors Description objectives and contact information Application process for employers to receive funding Internship opportunities application process for postsecondary graduates Interactive map of past internships across Canada You will not

- Attention sophomores Consider applying for the Student Climate Change Institute This is a great opportunity to connect with climate experts and take on projects that help address climate change Applications are due 928 httpstcoT0SHyuUbkL HCConservancy

- Are you an artist between the ages 1422 who wants to inspire social change Young New England artists have the chance at winning cash prizes of 2500 or 5000 for their artwork focused on climate change Submit to the TidalShiftAward today sparkleshttpstcoquf5WbOl6Lsparkles httpstcokxVJgqWhjb

- Weve put together some great STEM resources for parents who want to engage their kids on climate change do some learning and have some fun with these activities httpstcoMPKBqJt0aG Title Climate change and easy science experiments for kids Desciption At some point youre going to have to talk with your kids about climate change This can be intimidating so weve compiled some info on climate change for kids

- SF State launches a new interdisciplinary climate change certificate program this fall for students in any major httpstcogajrUVtFVJ httpstcoYVxiYQ52wX

Health

- A big number of children below the age of five are suffering from malnutrition coupled with extreme complications httpstcoFBG9ttoHko Title Over 100 children in Mandera exposed to droughtrelated diseases Desciption At least 106 children have already been admitted to hospitals across Mandera county

- The Environmental Protection Agency was never given agency to protect the environment Supreme Court rules httpstcouPYEmQLGLc Title Supreme Court rules for coalproducing states limits EPAs power to fight climate change Desciption Ruling in favor of coalproducing states Supreme Court says Congress not the EPA has the authority to make decisions on fighting climate change

- Breastfeeding breastfeedingmediumdarkskintoneproduces pushpinZero waste pushpinZero greenhouse gases and leaves pushpinZero water footprint seedling Its the healthiest and most natural way to feed babies Keep supporting breastfeeding for a cleaner healthier and more equal futureglobeshowingEuropeAfrica httpstcoHyXJO1Aj8Z

- Addressing Climate Change and improving livelihood opportunities for rural women through the production of reusable menstrual pads UNPeacebuilding Gambia UNFPATheGambia httpstcomNEAFYyPZX

- Via euronews Infertility heart failure and kidney disease How does climate change impact the human body httpstcoh3a7JwSScQ Title These are the 10 ways climate change affects our bodies every day Desciption We need the same urgency to treat climate change as when everyone jumped to combat the COVID19 pandemic Otherwise our health is due for a downward spiral in coming years

- One in five deaths every year results from unhealthy diets That is more than the number killed from smoking and armed conflict combined Pictet Nutrition funds Mayssa Al Midani tells MerrynSW why we need sustainable food systems Link here httpstcougzIvBeUKh httpstcozPmzAPfcFp

- backhandindexpointingdowndarkskintone Health problems During food shortages caused by climate change girls are more likely to go hungry and will often eat least and last leading to hunger and malnutrition httpstcoZcGNS4ROZL Title 5 ways climate change is disrupting girls lives Plan International Desciption These stories show how the inequalities experienced by marginalised girls and young women are amplified by the impacts of climate change

- Climate change affects the social and environmental determinants of health clean air safe drinking water sufficient food and secure shelter health healthcare healthy facts healthfact know more at httpstcoiTXON5Uiwn

- I wonder if climate change causes myocarditis and blood clots

- The pandemic has made things worse in addition to the rising costs of living in Aligarh Women are concerned about the high cost of food which they are unable to offset despite procuring supplies from ration shops writes aashi310 httpstcoGbf0cASadb

Rain

- RAIN FORECAST 7day rain forecast from the Weather Prediction Center Check radargt httpstcorJpa708eNn rain flood drought rainfall showers thunderstorm flooding httpstcoBaEKE6OOw7

- Storm Barra Severe weather warnings for wind and rain issued NewsEverything NorthernIreland httpstcozJohxg235r Title Online Shopping site in India Shop Online for Mobiles Books Watches Shoes and More Amazonin Desciption Amazonin Online Shopping India Buy mobiles laptops cameras books watches apparel shoes and eGift Cards Free Shipping Cash on Delivery Available

- Drenching showers and strong winds accompanied the weekends arrival of an atmospheric river a long and wide plume of moisture pulled in from the Pacific Ocean The National Weather Services Sacramento office warned of potentially historic rain httpstcoNvLHc5bq4e

- Greece Extreme Weather Warning Heavy Rains Thunderstorms Hurricane Winds httpstcoP2YB87bQam httpstcoVq6fZyNv7S

- RAIN FORECAST 7day rain forecast from the Weather Prediction Center Check radargt httpstcorJpa708eNn rain flood drought rainfall showers thunderstorm flooding httpstcomyR0wFpGE1

- UKWeather redcircle Thunderstorm warning issued as heavy rain set to soak England httpstcoN2yW0GF3Sr Title UK weather Thunderstorm warning issued as heavy rain set to soak England Desciption Downpour comes days after Britain recorded hottest temperature ever

- A lowpressure system brought extreme weather to Australia from swellbattered beaches to blizzardlike conditions httpstcopr0Xqt0PyU Title Watch Australian beaches battered by large swells Desciption A lowpressure system brought extreme weather to Australia from swellbattered beaches to blizzardlike conditions

- Flash flooding warnings after recordbreaking 40C heatwave Properties may flood httpstcocuzwzeEzIH Title Flash flooding warnings after recordbreaking 40C heatwave How climate change works Desciption FLASH FLOODING warnings have been issued for Wednesday after the UKs recordbreaking 40C heatwave with one analyst saying this is how climate change works

- Thunderstorms heading to Northeast following heat wave bringing drought relief httpstcoybW04G4tVQ qua usatoday Title Thunderstorms heading to Northeast following heat wave bringing drought relief Desciption The heat wave will send temperatures near 100 in the coming days in some cities while the storms will bring some needed drought relief

- Weather Buoys Ensure More Accurate Forecasts of Extreme Weather Events httpstcohPDi83TNyM

News

- Climate change The top environment news stories this week httpstcovLiW4IvjKs via wef Title Surprise trees early flowers and green football clubs Everything to know about the environment this week Desciption Top environment stories Research finds more than 73000 species of tree on Earth Plants in the UK are flowering a month earlier Cutting down on fossil fuelbased plastics is crucial to tackling climate change says EU environment chief
- Canadian onair weather personalities shifting tone amid worsening climate change National Newswatch httpstcorftrwyIYor httpstcoqFvWFkBCyX Title Canadian onair weather personalities shifting tone amid worsening climate change National Newswatch Desciption National Newswatch Canadas most comprehensive site for political news and views Make it a daily habit
- HAY Online News Brianna Sacks joins The Post as an extreme weather and natural disasters reporter The Washington Post httpstcodvNSWrHI88 httpstcolnwN8oTD6N Title Brianna Sacks joins The Post as an extreme weather and natural disasters reporter Desciption Sacks will explore how climate change is transforming the United States through violent storms intense heat widespread wildfires and other forms of extreme weather
- Climate change disclosures driving awareness and action among companies and investors TabbedNews News NewsToday Goodnews Breaking BreakingNews today story httpstcos8aNct9UdC
- fox vs msnbc lies about climate change vs lies about russia
- Oregon wildfires featured in new documentary Elemental Statesman Journal News httpstcoBSLdIQJe7x News BreakingNews Title Elemental film features Santiam Canyon fires looks to shift relationship with wildfire Desciption The documentary which includes footage from Oregon wildfires such as the Santiam Canyon and Eagle Creek fires is playing for a week at Salem Cinema
- Reuters a Pulitzer Prize finalist for feature photography on climate change httpstcoiMjKJQxlHt httpstcoflI0z7L2d2 Title Reuters A Pulitzer Prize Finalist For Feature Photography On Climate Change Desciption A general view can be seen from a damaged movie theater after a devastating tornado ripped through Mayfield Kentucky December 16 2021 REUTERSCheney Orr
- Extreme Weather Insiders httpstcocddbuRFjqq
- Live updates247 Breaking News httpstcoSgsI2yJkkg breakingnews Warming climate may boost Arctic virus spillover risk research shows httpstcoeIA4CcJmAM Title Live Update Breaking News 247 Financial Markets Desciption Breaking News affect finanical markets 247 searching all breaking news This channel sponsored by

Official InrexEA channel Inrexea provides trading robots and analysis service Title Warming climate could boost Arctic virus spillover risk research shows Desciption Its really unpredictable one research author said It can range from benign to an actual pandemic

- Biden Discusses Climate Change during Visit to florida Today shorts shortsvideo httpstco9hXX0PcaCV news politics conservative funny breakingnews alert httpstcoxwmTM9CJTg Title YouTube Desciption Enjoy the videos and music you love upload original content and share it all with friends family and the world on YouTube

Deaths

- Washington officials share video of four people and car connected to raging wildfire httpstcoKRvxHKvdY1 Title Washington officials share video of four people and car connected to raging wildfire Desciption Firefighters hiking on steep terrain with a 45pound backpack hand tools chainsaws and water throughout a 12hour shift state Department of Natural Resources says

- Heatwaveravaged areas in China now facing heavy rains prompting evacuations httpstcojQv74kS6RO Title Heatwaveravaged areas in China now facing heavy rains prompting evacuations National Globalnewsca Desciption Heavy rain in China was forecast for parts of Sichuan province and Chongqing city through at least Tuesday

- Flash floods and landslides set off by torrential rains swamped a southern Philippine province killing at least 42 people leaving 16 others missing and trapping some residents on their roofs officials said Friday httpstcoJtvJgHeQ7t Title At least 42 dead in floods landslides in south Philippines Desciption Flash floods and landslides set off by torrential rains swamped a southern Philippine province killing at least 42 people leaving 16 others missing and trappi

- My condolences are with the family amp friends of those whove lost their lives during this extreme weather event I also thank our emergency services and SES volunteers who are working around the clock to save lives amp protect properties

- Kentucky factory workers threatened with firing as tornado neared reports say The Independent news httpstcoQWr7zhn5nz

- At least 31 people have died 165 people are missing many more are feared to have died httpstcoJiO7qC3lH3

- On Monday August 1 search teams found two more bodies within the perimeter of the McKinney wildfire in Northern California httpstcoXXBk7NgazC Title Wildfire death toll rises to 4 in Northern California Desciption On Monday August 1 search

teams found two more bodies within the perimeter of the McKinney wildfire in Northern California

- On Friday the India Meteorological Department issued a fiveday severe weather warning alert for multiple parts of the country as temperatures in some areas reach more than 113 degrees httpstco9sbaoqHNPY
- Wind and rain from the storm caused downed trees and power lines as well as flooded roads in some regions of the state httpstco3SUBcD2urM
- Heavy rains fell across Taiwan on Sunday alleviating the drought in some areas and causing flooding in Changhua County according to the Central Weather Bureau CWB httpstcoo51mbRYiqJ

Investment

- cfauk has launched a qualification designed to help the investment management industry understand the implications of climate change on investments by RLawther94 FinancialAdvisers WealthManagement httpstcoXrDDSHiWHE
- Would a prudent fiduciary make comp disclosures based on such unlikely events and negative investment returns its top priorityresearch reflects that some ESG funds have underperformed A vIolation of fiduciary responsibility Shareholder lawsuits httpstcos7FgtNY2I2
- Most of the information the SEC wants companies to disclose is irrelevant to financial performance It would also expose companies to progressive bullying and classaction lawsuits httpstcoeHrkN4a9mj Title Opinion Gary Gensler Stonewalls Congress Desciption The SEC chief refuses to answer questions about his climate rule
- Wall Street is trying to fool investors into thinking they can get rich and save the planet at the same time pareene writes httpstcozfkKL0MUmz Title ClimateFriendly Investment Funds Are a Scam Desciption Wall Street is trying to fool investors into thinking they can get rich and save the planet at the same time
- Insurance companies really should seek funds from fossil fuel companies to cover climate related damages httpstco5O2ZyWsPUO
- Big US banks have utterly failed to protect their shareholders longterm interests as they renege on their netzero commitments httpstcoBs1b0PADH3 Title Investors Must Hold Banks to Their Word on Climate Change Desciption The biggest financiers of fossil fuels face some tough questions this shareholder season
- GWSO Share Repurchase Announcement Causes Premarket Buying Pressure httpstcoNGIOGofwp0

- Italys De Nora bets on cornerstone investors to defy IPO drought httpstcooeeI3qtnnj Title Italys De Nora bets on cornerstone investors to defy IPO drought Desciption Published by Reuters UK By Francesca Landini and Lucy Raitano MILAN Reuters Italys Industrie De Nora is counting on cornerstone investors to defy volatility that has inhibited several initial
- Dividends from British companies crashed last year and could fall even further in 2021 but the best investment trusts can offer investors sanctuary from the cuts httpstcoIRI0T2Et3Y
- NEWS QU has millions in hedge fund investments One hedge fund puts millions into fossil fuel industries every year link httpstcorMs5gr07M6 via stevemac2017 httpstcobEhhy71Z8F

Human cost

- BONFIRES As well as potentially causing a nuisance bonfires can produce greenhouse gases such as carbon dioxide which add to global warming Bonfires can also produce other poisonous gases and fine particles which can affect human health
- There will be drought throughout the United States lots of people will suffer because of food shortages and escalating food prices
- Floods and droughts are some of the most tangible and devastating consequences of the climate crisis httpstco1Jhj8fntI7
- extreme weather patterns and blackouts caused as a result of such are always terrifying especially since the people who are affected most by them are typically those who are already most vulnerable disabled poor homeless and marginalised people are often at significant risk
- Very Concerning Sea level rise could threaten hundreds of toxic sites in California httpstcoOfsQHg5FOO
- Effects Climate Change causes serious problems like droughts floods extinction of animals high temperatures rising sea levels etc httpstcoubX4oE8S7r
- Fires to floods Extreme weather is occurring worldwide httpstcoIdoUhsdvp2
- Building in DisasterProne Areas Not More Extreme Weather Causes Rising Losses httpstcok6KMZRps6C
- RT BrookingsInst Dramatic storms wildfires and floods often generate headlines but more subtle persistent changes in the environment are also creating health and safety hazards across the United States httpstcozp0Yef92RC Title Sea level rise from climate change is threatening home septic systems and public health

Desciption Dramatic storms wildfires floods and similar events draw the most public attention as examples of how climate change threatens human lives and homes But more subtle persistent changes in the e

- More people are moving into dangerous areas as climate change is making weather disasters stronger and more frequent httpstcoEMa2pIf0CL

Projections

- The plan could allow for emissions to keep growing through 2025 httpstco0CwBsBOBVX via voxdotcom lilipike
- httpstcotIbFjtofPl 30year global projections More unprecedented droughts ahead
- Most of the new units will be available next year but some could be up as early as this fall httpstcoIGUR20jKYE Title Younger generations could be the key to reaching older Republicans on climate change Desciption Utah Tech hosted a conservative climate change panel that focused on the way to bridge the gap between younger and older generations on the issue
- Earth could cross the global warming threshold as soon as 2027 httpstcoVtO4jMOrig httpstcobpH9z7XFz7
- For a 67 chance of limiting global warming to 15C we would have to reach netzero by 2030
- There is a 5050 chance globalwarming will exceed 15C before 2025 WMO says httpstcoRbhcdakzmF Title There is 5050 chance global warming will exceed 15C before 2025 WMO says Desciption The WMO warns of 5050 chance that global warming will exceed 15 degrees within next five years
- 2020 was a preview of hotter years to come httpstco4uEiJZms7W
- Under all scenarios examined Earth is likely to reach the crucial 15 warming limit in the early 2030s httpstcoxrDjsArY7Q
- We could be seeing some good news for the fall outlook httpstcoJlayQnKOMJ
- Climate change will get worse in 2022 But it wont be the end httpstco7oxDS12WBF Title Climate change will get worse in 2022 But it wont be the end Desciption The best we can hope for is incremental progress two steps forward one step back a string of little victories

Links/Promo

- Check out this cartoon httpstcosjx1vDnk6c via TheWeek Title Check out this cartoon Desciption Editorial Cartoons from The Week

- Check out this cartoon httpstco1GTmMWFHAJ via TheWeek
- Extreme Weather Warning Please click the link below for our update httpstcoUIYbhAtLRT httpstcoBa1v9SzTfK Title Extreme weather Update Desciption Dear Parent Carer Further to the communication sent out last week please be advised that students can continue to wear uniform with no blaze
- Climate summit If anyone is struggling to get in to the webinar please try again and make sure you are logged in to your zoom account use link on this page httpstcogHGQAWSE4k netzerosheffield
- You can listen to and subscribe to the podcast on Apple Podcasts here httpstcoVgkoxl5u9q Title We havent faced anything like climate change before Gaia Vince Desciption In this episode Gaia Vince joins Krishnan to talk about her new book Nomad Century in which she takes a look at how migration could be the solution to the climate crisis
- Get my art printed on awesome products Support me at Redbubble RBandME httpstcok3UhxySc3p findyourthing redbubble Title Climate Change Is Real Act Now Tote Bag by Peter Baker Desciption The time to act is now Climate change is real The sea ice is melting polar bears wont survive eating coconuts ENJOY Millions of unique designs by independent artists Find your thing
- fossilfuelfreefriday What do you think about this announcement following COP26 Click this link to learn more httpstcoia7YB8689K httpstco1f3mfmCV5D
- Check out this article I found on Knewz httpstcopywaqXAF5Z Get more Knewz iOS app httpstco6S79hVObml Android app httpstcoZ3n1eS957s Online httpstcoT301uhWrEG
- Check out this cartoon httpstcop6TPMkc3mv via TheWeek
- Get my art printed on awesome products Support me at Redbubble RBandME httpstcoXiJBJJo1LC findyourthing redbubble Title Climate Change Code Red Tote Bag by Peter Baker Desciption Climate change code red the time to act is now before its too late ENJOY your earth dont abuse it Millions of unique designs by independent artists Find your thing

B Lobbying

The names of all lobby clusters along with the domains of their members are given in Table B.1.

Lobby Cluster	**Lobby Domains**
Agriculture Interest Groups.-0	agroecology-europe.org, animaltaskforce.eu, beeflambnz.com, beesfordevelopment.org, celep.info, cesfac.es, ciwf.eu, coleacp.org, dairyuk.org, dvtier-nahrung.de, efncp.org, iatp.org, nevedi.nl, pollinis.org, risefoundation.eu, save-foundation.net, tporganics.eu, uecbv.eu, wsrw.org
Entertainment - 1	aereurope.org, audiogest.pt, baseorg.uk, composeralliance.org, culture-media.eu, emc-imc.org, europeanfilmagencies.eu, fim-musicians.org, ietm.org, ifpi.org, impalamusic.org, impforum.org, irma.ie, mmta.co.uk, scpp.fr, ukmusic.org, weee-forum.org, worlddab.org
Human rights focused groups.-2	alliancevita.org, crd.org, ebco-beoc.org, ecchr.eu, ejfoundation.org, frankbold.org, icj.org, lastradainternational.org, liberties.eu, odfoundation.eu, oidel.org, panoptykon.org, rsf.org, saamicouncil.net, savetibet.org, silc.se
Social Justice - 3	aefjn.org, antislavery.org, attac.at, ceji.org, cidse.org, emmaus-europe.org, enar-eu.org, equineteurope.org, germanwatch.org, idsn.org, oijj.org, oxfammagasinsdumonde.be, oxfamwereldwinkels.be, profesionalesetica.org, s2bnetwork.org, socialplatform.org, solidar.org, wecf.org, weed-online.org

Business (small group) - 4	financelatvia.eu, ispa.at, voeb.de, wvmetalle.de, zpbsp.com
European industry interest groups.-5	agw.org.au, apiccaps.pt, assarmatori.eu, cec-footwearindustry.eu, ceev.eu, cefic.org, cerameunie.eu, cov.nl, donboscointernational.eu, ebca-europe.org, ecovin.de, efjewellery.eu, eicf.org, eucolait.eu, euromines.org, federlegnoarredo.it, hotrec.eu, irish-exporters.ie, izbamleka.pl, liquidgaseurope.eu, livsmedelsforetagen.se, metsastajaliitto.fi, nzo.nl, puutuoteteollisuus.fi, scotch-whisky.org.uk, sustain-ablefur.com
Medical advocacy groups.-6	amdr.org, beam-alliance.eu, chiropractic-ecu.org, deutsche-diabetes-gesellschaft.de, eadv.org, eanm.org, ecetoc.org, efcni.org, eggvp.org, elpa.eu, en-docrine.org, ese-hormones.org, esot.org, essm.org, ewma.org, homeopathyeurope.org, iadr.org, pandem-icactionnetwork.org, pptaglobal.org, tballiance.org, tbvi.eu
Standards interest groups.-7	anec.eu, cepis.org, ecma-international.org, efrag.org, floricode.com, gs1.eu, gs1.org, iabslovakia.sk, iec.ch, nen.nl, pharmacyregulation.org, revisorforeningen.no, sbs-sme.eu, tic-council.org

Business - 8	afep.com, afera.com, anie.it, arc2020.eu, barcouncil.org.uk, bevh.org, boersenverein.de, bpf.org.uk, cbi.org.uk, cc.lu, ccci.org.cy, cebre.cz, ceoe.es, cip.org.pt, cnipmmr.ro, concordeurope.org, coopseurope.coop, da.dk, dafne-online.eu, danskerhverv.dk, eap-csf.eu, earto.eu, eboworldwide.eu, einzelhandel.de, elf-fae.eu, enterprisealliance.eu, esba-europe.org, eurochamvn.org, eurocities.eu, eurocrowd.org, fedil.lu, finnwatch.org, fsb.org.uk, hup.hr, iab.org.pl, ibec.ie, iccgermany.de, iccwbo.org, ifglobal.org, ila-lead.org, independentretaileurope.eu, integratedreporting.org, leasingverband.de, norskindustri.no, nvo.lv, oeb.org.cy, pisil.pl, remancouncil.eu, sa.is, seldia.eu, smeeurope.eu, spcr.cz, spirituosenverband.de, svenskhandel.se, svensktnaringsliv.se, taxjustice.net, tei.org, thefactcoalition.org, tusiad.org, vbo-feb.be, vending-europe.eu, vno-ncw.nl, vnp.nl, wettbewerbszentrale.de, wfanet.org, wise-europa.eu, zia-deutschland.de, zpp.net.pl
Business and industry interest groups.-9	accessibletourism.org, aquatt.ie, bingo-brussels.eu, britishchambers.org.uk, bvoed.de, ceeman.org, centr.org, eespa.eu, efci.eu, epaca.org, erarental.org, eubfe.eu, eurid.eu, europeanfamilybusinesses.eu, eurowindoor.eu, federgon.be, ficil.lv, geode-eu.org, keidanren.or.jp, rehva.eu, servicealliance.eu, shrm.org, stm-assoc.org, thesynergist.org
Energy Industry Advocacy.-10	afgnv.org, co2value.eu, energigas.se, entsog.eu, europeanbiogas.eu, euturbines.eu, gerg.eu, hydrogeneurope.eu, industriegaseverband.de, mew-verband.de, ngva.eu, oilgasdenmark.dk, sea-lng.org, slocat.net, ukoog.org.uk, ukpia.com
Energy and trade interests.-11	aib-net.org, entsoe.eu, europex.org, smarten.eu, wuwm.org
Sustainable Development Groups.-12	alliance2015.org, antaisce.org, arij.org, asvis.it, ccifer.ro, milieudefensie.nl, sdgwatcheurope.org, zero.ong
Cancer advocacy groups.-13	breastcanceruk.org.uk, cancer.dk, cancernurse.eu, digestivecancers.eu, ebmt.org, eortc.org, essoweb.org, estro.org, europeancancerleagues.org, fondationarcad.org, komoptegenkanker.be, kwf.nl

Technology advocacy groups.-14	bitsoffreedom.nl, blockchain4europe.eu, cdt.org, digitalegesellschaft.de, dinl.nl, ecommerce-europe.eu, espi.or.at, fedma.org, ficom.fi, ftthcouncil.eu, gdd.de, gp-digital.org, homodigitalis.gr, i2coalition.com, iabeurope.eu, informatics-europe.org, internetforum.eu, internetsociety.org, ipc.org, isfe.eu, itic.org, openmedia.org, privacyinternational.org, sos-save-our-spectrum.org, techuk.org, teknikforetagen.se, thefuturesociety.org
Entrepreneurship interest groups.-15	euclidnetwork.eu, ied.eu, mkb.nl, youthproaktiv.org, yrittajat.fi
Advocacy for Democracy and Good Governance.-16	assemblea.cat, avaaz.org, batory.org.pl, ceceurope.org, clubmadrid.org, democracy-international.org, enop.eu, epd.eu, freiheit.org, gong.hr, laicite.be, ndi.org, oziveni.cz, sol-asso.fr, tpnonline.org, transparency.nl
Advocacy for libraries.-17	cenl.org, ebib.pl, eskillsassociation.eu, libereurope.eu, publiclibraries2030.eu, publishingireland.com, ucl.ac.uk
Industry interest groups.-18	amaplast.org, anima.it, bavc.de, bdia.org.uk, cembureau.eu, cosmeticseurope.eu, ermco.eu, fem-eur.com, fepa-abrasives.org, ikem.se, ima-europe.eu, intergraf.eu, isopa.org, modernbuildingalliance.eu, ofice-men.com, plasticsconverters.eu, plasticseurope.org, pu-europe.eu, teppfa.eu, unic.it, vnci.nl, wdk.de
Humanitarian Aid Groups.-19	care.at, eu-cord.org, heks.ch, ifrc.org, msf.org, realityofaid.org, sboverseas.org, sea-watch.org, voiceeu.org
Financial and legal protection.-20	cifar.eu, csiworld.org, e-ma.org, eccbelgium.be, elen.ngo, feat-alliance.org, fidoalliance.org, globalplatform.org, gvg.org, insol-europe.org, promarca-spain.com, tapaemea.org, trustindigitallife.eu, verbraucherzentrale-bawue.de, whistleblowingnetwork.org
Finance and Investment Groups.-21	aifi.it, aima.org, alfi.lu, asifma.org, bettingandgamingcouncil.com, capitalscoalition.org, crefceurope.org, eban.org, efama.org, eltia.eu, european-lotteries.org, icmagroup.org, investeurope.eu, ipf.org.uk, irishfunds.ie, paaomasijoittajat.fi, world-exchanges.org

Diverse Health Interests.-22	accesstomedicinefoundation.org, aemh.org, aerztederwelt.org, afew.org, cam-europe.eu, eapaediatrics.eu, echalliance.com, efort.org, ehff.eu, epfweb.org, ersnet.org, eupatientaccess.eu, eupha.org, eurodiaconia.org, eurohealth.ie, eurohealthnet.eu, europeactive.eu, europsy.net, fondation-merieux.org, gatesfoundation.org, girp.eu, i-hd.eu, ifmsa.org, ihe-europe.net, ippfen.org, maphm.org, path.org, pcdeurope.org, pvcmed.org, snomed.org
Waste and Resource Management.-23	asegre.com, biokierto.fi, cleaneuropenetwork.eu, compost.it, compostnetwork.info, esauk.org, eswet.eu, eurofoodbank.org, fead.be, feedbackglobal.org, fnade.org, matvett.no, smartwasteportugal.com, solaal.org, wastematters.eu, zerowasteeurope.eu
Rural and urban development.-24	apdes.pt, deutscher-verband.org, elard.eu, euromontana.org, europeanlandowners.org, gaq.be, reseaupwdr.be, sspa-network.eu, sverigesallmannytta.se
International Trade Interest Groups.-25	aicebiz.com, amcham.de, amcham.fi, amcham.ie, amchameu.eu, amfori.org, apexbrasil.com.br, atahq.org, auma.de, bga.de, bimco.org, britishirishchamber.com, clubexportadores.org, commercequitable.org, crossborder.ie, diplomats.pl, eaccny.com, ebc-jp.com, fairtrade.net, feex.org, globalshippersforum.com, globsec.org, hgk.hr, ics-shipping.org, lngallies.com, publiceye.ch, tracit.org, transatlanticbusiness.org, uschamber.com, weforum.org, wfto-europe.org
Animal welfare advocacy.-26	animalhealtheurope.eu, animals-angels.de, animaltransportationassociation.org, bft-online.de, bluecross.org.uk, crueltyfreeeurope.org, crueltyfreeinternational.org, djurensratt.se, djurskyddet.se, dyrenesbeskyttelse.dk, furfreealliance.com, ivsa.org, petcoreeurope.org, rspca.org.uk, vgt.at, welfarm.fr, worldanimalprotection.org

Energy advocacy groups.-27	bee-ev.de, caneurope.org, cewep.eu, cgoa.cz, communityenergyengland.org, danskenergi.dk, dwv-info.de, ease-storage.eu, efet.org, ehi.eu, energi-foretagen.se, energy-uk.org.uk, energycoalition.eu, energysavingtrust.org.uk, eurogas.org, febeg.be, fundacionrenovables.org, kernd.de, marcogaz.org, nuclear-transparency-watch.eu, oftec.org, r-e-a.net, ren21.net, solarimpulse.com, wind-energie.de, world-nuclear.org
Business and legal advocacy.-28	biac.org, cfdverband.de, cryptovalues.eu, dnotv.de, drc.ngo, ecsda.eu, enaat.org, eocc.nu, etno.eu, eu-lita.eu, eurofi.net, evia.org.uk, fese.eu, globalpol-icy.org, halotrust.org, icoca.ch, indicam.it, permits-foundation.com, radiocentre.org, ruzsr.sk, spir.cz, vd-eh.de, work-with-perpetrators.eu, worldjusticepro-ject.org
Advocacy for various sports.-29	cttc.ie, cyclingindustries.com, dualcareer.eu, egba.eu, essna.com, eurolympic.org, europeanleagues.com, fesi-sport.org, ibia.bet, ifhaonline.org, isca-web.org, is-fsports.org, mission89.org, paralympic.org, resul.fi, sroc.info, tafisa.org, theicss.org, wada-ama.org
Nature conservation interest groups.-30	arocha.org, awf.org, bornfree.org.uk, buglife.org.uk, businessfornature.org, butterfly-conservation.org, cipra.org, conservation.org, eaam.org, edf.org, ethicalbiotrade.org, euronatur.org, europarc.org, face.eu, iaf.org, iucn.nl, mammiferi.org, nature.org, nwf.org, prowildlife.de, thehabitatfoundation.org, theperfectworld.com, traffic.org, tropenbos.org, umweltdachverband.at, wcs.org, wildlifejustice.org, wwf.be, wwf.fi, wwf.fr, wwf.it, wwf.nl
Heritage-oriented interest groups.-31	europanostra.org, exarc.net, frh-europe.org, fun-dacionacm.org, icom.museum, michael-culture.eu, schuman-seura.fi, socantscot.org
HIV/AIDS advocacy and support.-32	coalitionplus.org, eatg.org, hivjustice.net, ipopi.org
Children's rights advocacy.-33	childcircle.eu, childfundalliance.org, eurochild.org, hopeandhomes.org, makemothersmatter.org, miss-ingchildreneurope.eu, oco.ie, sos-childrensvillages.org, supportkind.org, terredeshommes.org

Aquatic industry interest groups.-34	aac-europe.org, anfaco.es, britishtrout.co.uk, eatip.eu, effop.org, feap.info, ornamentalfish.org, seafoodalliance.org, sustainableeelgroup.org
Sustainability - 35	92grp.dk, aise.eu, bef.lv, bothends.org, changingmarkets.org, ciel.org, ecopreneur.eu, ecostandard.org, eeb.org, entretantos.org, environmentalpillar.ie, eu-umweltbuero.at, fidra.org.uk, foemalta.org, forumue.de, mightyearth.org, noharm-europe.org, seechangenetwork.org, umanotera.org, walk21.com, wbcsd.org
Emerging technology interest groups.-36	arpas.uk, borealis.aero, broadcast-networks.eu, earsc.org, eata.be, edsoforsmartgrids.eu, encs.eu, errin.eu, ewia.org, gigaeurope.eu, iapa.org, icann.org, lora-alliance.org, maas-alliance.eu, oascities.org, vleva.eu
Disability advocacy group.-37	enil.eu, epr.eu, euroblind.org, euse.org, iddconsortium.net, medaxes.be, once.es, specialolympics.org
Marine Conservation Groups.-38	asosalimar.com, bluemarinefoundation.com, ccb.se, eurogoos.eu, gceocean.no, jpi-oceans.eu, mcsuk.org, mundusmaris.org, oceana.org, oceancare.org, oceancouncil.org, oceanoazulfoundation.org, panda.org, savethehighseas.org, sciaena.org, seas-at-risk.org, sharkproject.org, sharktrust.org, surfrider.eu,
Set of Bio-based Advocacy Groups.-39	appa.es, asebio.com, assobioplastiche.org, bbia.org.uk, biconsortium.eu, ebb-eu.org, epure.org, etipbioenergy.eu, europabio.org, european-bioplastics.org, fems-microbiology.org, hollandbio.nl, ibma-global.org, irbea.org, norman-network.net, pharmabiotic.org, sbp-cert.org, svebio.se
Women's rights advocacy.-40	afaemme.org, bpw-europe.org, endfgm.eu, sexworkeurope.org, vrouwenrecht.nl, wave-network.org, women-politicalleaders.org

Food and agriculture interest groups.-41	albert-schweitzer-stiftung.de, asedas.org, avec-poultry.eu, bogk.org, countryside-alliance.org, cpc-ccp.com, ensa-eu.org, euoilseed.org, eurofir.org, euroveg.eu, feder.bio, fnli.nl, globalharmonization.net, losp.lv, milchindustrie.de, nifda.co.uk, nmpf.org, pan-europe.info, peanutsusa.com, plantbasedfoodalliance.eu, safefoodadvocacy.eu, scottishsalmon.co.uk, tappcoalition.eu, unionfleurs.org, vegansociety.com, voicenetwork.eu, wervel.be, zemniekusaeima.lv, zscr.cz
Human rights interest groups.-42	actsa.org, amnesty.eu, aprnet.org, btselem.org, caj.org.uk, cnapd.be, cospe.org, docip.org, eccpalestine.org, ecdhr.org, ecnl.org, ennhri.org, ethicaltrade.org, forum-asia.org, freedomofconscience.eu, helsinki.hu, hfhr.pl, hrw.org, hrwf.eu, humanistfederation.eu, humanrightshouse.org, ibonintern ational.org, ictj.org, ilga-europe.org, indexoncensorship.org, iphronline.org, ishr.ch, ituc-africa.org, justiceandpeace.nl, minorityrights.org, nhc.nl, privacyfirst.nl, rainbowrose.eu, reproductiverights.org, tgeu.org, wideplus.org, wo-men.nl, womenlobby.org
Diverse science interest groups.-43	citizen-science.net, ecsite.eu, efmi.org, egu.eu, elifesciences.org, eu-life.eu, eurogeosurveys.org, euronuclear.org, eurotech-universities.eu, eusea.info, fnp.org.pl, isscr.org, nanotechia.org, stem.cz, tour4eu.eu, urheberrechtsbuendnis.de, vdgh.de, volkswagenstiftung.de, worldfuturecouncil.org, yacadeuro.org
Safety advocacy and protection.-44	asb.de, brandskyddsforeningen.se, ctif.org, eena.org, efus.eu, electricalsafetyfirst.org.uk, euralarm.org, eurosprinkler.org, gndr.org, ime.org, iosh.com, mwfai.org, origin-gi.com, psc-europe.eu

Renewable Energy - 45	airbornewindeurope.org, californiahydrogen.org, ceep.be, ceer.eu, clasp.ngo, deneff.org, door.hr, ee-isac.eu, efiees.eu, egec.org, energinorge.no, energy-cities.eu, energy-transitions.org, energynetworks.org, equilibredesenergies.org, eref-europe.org, esmig.eu, estelasolar.org, euroace.org, fire-italia.org, geoexchange.ro, greenreality.fi, hydrogencouncil.com, ifieceurope.org, inforse.org, oceanenergy-europe.eu, psew.pl, rescoop.eu, ruralelec.org, rurener.eu, seforall.org, solarpowereurope.org, theade.co.uk, ve.dk, windeurope.org, zeroemissionsplatform.eu
Digital Rights Groups.-46	accessnow.org, article19.org, bildkunst.de, cisac.org, communia-association.org, digitalcourage.de, edri.org, eema.org, eff.org, epicenter.works, globalnetworkinitiative.org, ifrro.org, internews.org, itpol.dk, iwf.org.uk, urheber.info
Prevention and advocacy groups.-47	alcohol-focus-scotland.org.uk, alcoholireland.ie, aldp.ie, alliancechronicdiseases.org, cepi.net, eapcct.org, easo.org, ehnheart.org, eurocare.org, euspr.org, movendi.ngo, ntakk.lt, shaap.org.uk, woncaeurope.org
Pharmaceutical and Chemical Advocacy - 48	affordablemedicines.eu, chemtrust.org, efspi.org, eipg.eu, emvo-medicines.eu, epsa-online.org, eucope.org, eurad.net, europeantissue.com, eurovape.eu, fecc.org, federchimica.it, haiweb.org, ifi.hr, ikw.org, inpud.net, medicinesforeurope.com, medicinesforireland.ie, methanol.org, natrue.org, pgeu.eu, producencilekow.pl, progenerika.de, sfee.gr
Education Interest Groups - 49	actionuni.ch, atee.education, coimbra-group.eu, eadtu.eu, eden-online.org, efvet.org, enqa.eu, euchems.eu, eufic.org, eun.org, eunis.org, eurashe.eu, evbb.eu, evta.eu, iaapsy.org, kultur-life.de, learningandwork.org.uk

Manufacturing - 50	aia-aerospace.org, aijn.eu, asd-europe.org, aspapel.es, bdsv.eu, beerandpub.com, cece.eu, cecip.eu, ceereal.eu, cefs.org, cirfs.org, edana.org, effpa.eu, egeaassociation.eu, egmf.org, emeca.eu, epmf.be, eurometaux.eu, europacable.eu, europanels.org, europeansunlight.eu, federmeccanica.it, fediol.eu, fefac.eu, fefco.org, feve.org, finat.com, fooddrinkeurope.eu, glassforeurope.com, internationaltin.org, kalkzandsteen.nl, madridaerospace.es, makeuk.org, medialitto.fi, orgalim.eu, sea.org.uk, sugarrefineries.eu, unesda.eu, vereniging-ion.nl
Social advocacy groups.-51	agefriendlyeurope.org, b-b-e.de, balkancsd.net, bankofcyprus.com.cy, bertelsmann-stiftung.de, caritascoimbra.pt, civicus.org, esn-eu.org, feantsa.org, regionsunies-fogar.org, solidaritynow.org, uclga.org, wheel.ie
Corporate accountability and governance.-52	csreurope.org, dif.fi, ecgi.global, efc.be, eumedion.nl, financialtransparency.org, institute.global, iod.com, lobbycontrol.de, pwyp.org, swedwatch.org
Peace and conflict advocacy.-53	cooperationireland.org, ec4i.org, eplo.org, euracnetwork.org, globalwitness.org, icanw.org, kofiannanfoundation.org, paxforpeace.nl, rondine.org, sfcg.org
Recycling industry interest groups.-54	deutsche-phosphor-plattform.de, eera-recyclers.com, egaranet.org, etira.org, euric-aisbl.eu, ewaba.eu, rreuse.org, water-reuse-europe.org, worldloop.org
Maritime interest groups.-55	ecsa.eu, empa-pilots.eu, eumos.eu, eurotugowners.com, feport.eu, inlandports.eu, ivr-eu.com, kvnr.nl, maritimetechnology.nl, medcruise.com, pianc.org, pole-mer-bretagne-atlantique.com, shipbreakingplatform.org, ukchamberofshipping.com, worldshipping.org
Environmental advocacy interests.-56	airclim.org, chemsec.org, cittadiniperlaria.org, eiainternational.org, env-health.org, gahp.net, hejsupport.org, ifeh.org, justiceandenvironment.org, notreaffaireatous.org, pollens.fr, pureearth.org, southernenvironment.org, uecna.eu, wwf.de, zazemiata.org

Transportation interest groups.-57	a4e.eu, aef.org.uk, anat.ro, aoa.org.uk, bdo.org, bgl-ev.de, citainsp.org, clecat.org, confetra.com, cpt-uk.org, debatingmobility.eu, ecgassociation.eu, ec-sla.eu, ectri.org, eimrail.org, eraa.org, erfarail.eu, eutraveltech.eu, fntr.fr, ftai.ie, intertanko.com, its-mobility.de, leia.co.uk, mafex.es, railworking-group.org, tiaca.org, tln.nl, uic.org, vdik.de, wttc.org
Textile and labor advocacy.-58	atevalinforma.com, cleanclothes.org, dmogt.dk, eura-tex.eu, textile.fr
Interests in food industry.-59	barillacfn.com, beveragecarton.eu, choicespro-gramme.org, effca.org, ehpm.org, eitfood.eu, esasnacks.eu, eseb.org, eucofel.eu, foodandwatereu-rope.org, ilsi.eu, ipiff.org, iseki-food.net, medicalnu-tritionindustry.com, oenoppia.com, slowfood.com, specialisednutritioneurope.eu, sweeteners.org, tdmr-europe.com
Insurance and advocacy groups.-60	abi.org.uk, biba.org.uk, bundderversicherten.de, ffa-assurance.fr, forsikringogpension.dk, insuranceeu-rope.eu, iumi.com, pkv.de, reinsurance.org, svenskfor-sakring.se, svv.ch, verzekeraars.nl, voev.de
Health advocacy groups.-61	bhf.org.uk, braincouncil.eu, dystonia-europe.org, ean.org, efanet.org, efna.net, ehc.eu, emhalliance.org, emsp.org, eu.com, eunaapa.org, eurohuntington.org, itf.si, parkinsons.org.uk, thalassaemia.org.cy, vsop.nl
Entrepreneurial Interest Groups.-62	cbba-europe.eu, coadec.com, europeanstartupnet-work.eu, familienunternehmen.de, investmentmigra-tion.org, unitee.eu
Industry advocacy groups.-63	aluinfo.de, applia-europe.eu, constructionprod-ucts.org.uk, ebc-construction.eu, eiha.org, eurofer.eu, eurogypsum.org, evia.eu, fuelseurope.eu, iadc.org, iva.de, jernkontoret.se, kaivosteollisuus.fi, london-miningnetwork.org, nam.org, plasticsindustry.org, stahl-online.de, ufip.fr, zvei.org
Agricultural interest groups.-64	agindustries.org.uk, agricord.org, ailimpo.com, bv-agrar.de, ceia3.es, cema-agri.org, ciaracec.com.ar, co-ceral.com, cofalec.com, confagri.pt, eurofoiegras.com, europatat.eu, europeanfoodforum.eu, frucom.eu, graan.com, mpoc.org.my, nefyto.nl, raiffeisen.de, sos-faim.be, usmef.org, ussec.org

Digital and ICT interest groups.-65	adan.eu, all-digital.org, apdsi.pt, bdva.eu, bdzv.de, ccianet.org, cispe.cloud, digitales.es, digitaleurope.org, dlearn.eu, ecis.eu, eco.de, ectaportal.com, edpia.eu, ega.ee, ehtel.eu, eudca.org, eurodig.org, francedigitale.org, tice.pt, zeker-online.nl
Road safety and transportation advocacy.-66	5gaa.org, alpeninitiative.ch, asecap.com, astazero.com, eapa.org, earpa.eu, eiturbanmobility.eu, erf.be, ertrac.org, esporg.eu, eupave.eu, eurorap.org, fevr.org, irfnet.ch, pedestrians-int.org, theicct.org, transfrigoroute.eu
Human rights advocacy.-67	bnaibritheurope.org, ecre.org, ergonetwork.org, errc.org, forumrefugies.org, hias.org, icmc.net, ijl.org, interwencjaprawna.pl, jrseurope.org, migrantwomennetwork.org, nelfa.org, npld.eu, nrc.no, picum.org, refugee-rights.eu, rescue.org, romeurope.org, siriusmigrationeducation.org, tampep.eu
Plant-related interest groups.-68	aifm.org, aiph.org, aiprom.ro, arche-noah.at, beelife.eu, biostimulants.eu, croplife.org, ecofi.info, fertilizerseurope.com, glastuinbouwnederland.nl, iceers.org
Advocacy for marginalized groups.-69	age-platform.eu, alzheimer-europe.org, amberalert.eu, autismeurope.org, bagfw.de, bagso.de, cbm.org, cermi.es, coface-eu.org, coteceurope.eu, dianova.org, driadvocacy.org, easpd.eu, edbn.org, edf-feph.org, efhoh.org, eud.eu, eufami.org, fafce.org, funktionsratt.se, gamian.eu, gezinsbond.be, harmreductioneurasia.org, horatio-web.eu, hri.global, inclusioneurope.eu, irct.org, light-for-the-world.org, mhesme.org, validity.ngo
Water management interest groups.-70	aquapublica.eu, asersagua.es, bdew.de, coalitioneau.org, euraqua.org, eureau.org, ewa-online.eu, iah.org, igwp.org.pl, inbo-news.org, inlandnavigation.eu, wateraid.org, womenforwater.org

Pro-EU interest groups.-71	abe-eba.eu, alter-eu.org, brill-luxembourg.org, c4ep.eu, civilsocietyeurope.eu, ehfg.org, eulat-network.org, eurochambres.eu, europa-union.de, europeanconstitution.eu, europeanmovement.eu, europeanmovement.ie, europeanpaymentscouncil.eu, federalists.eu, iep-berlin.de, iucn.org, kent.ac.uk, neweuropeans.net, rewildingeurope.com, samaritan-international.eu, sbra.be, united-europe.eu, volteuropa.org
Education and advocacy groups.-72	brot-fuer-die-welt.de, dsfnet.dk, eaea.org, earlall.eu, emsa-europe.eu, eucen.eu, euromil.org, eusalt.com, iau-aiu.net, iea.nl, lllplatform.eu, londonhigher.ac.uk, neth-er.eu, stiftung-mercator.de, the-guild.eu, unifi.fi
Forest advocacy groups.-73	bauernverband.de, cepf-eu.org, eustafor.eu, fern.org, forestindustries.se, forestplatform.org, global-canopy.org, iflaeurope.eu, tropicalforestalliance.org
Circular economy advocacy groups.-74	acrplus.org, circularchange.com, ecorec.gr, ellen-macarthurfoundation.org, hollandcircularhotspot.nl, institut-economie-circulaire.fr, rediscoverycentre.ie, ytpliitto.fi
Health advocacy groups.-75	acmedsci.ac.uk, acrohealth.org, aesgp.eu, aides.org, aim-mutual.org, alliancerm.org, amrc.org.uk, anhinternational.org, bma.org.uk, eaasm.eu, eapcnet.eu, ecpc.org, ekha.eu, enrf.eu, epha.org, erwcpt.eu, euipff.org, eu-patient.eu, eular.org, eumca.org, euprimarycare.org, euradia.org, europadonna.org, europeanpainfederation.eu, eurordis.org, fertilityeurope.eu, forestonline.eu, france-assos-sante.org, global-sepsis-alliance.org, inspire2live.org, ivaa.info, lungcancereurope.eu, naery.fi, nhsconfed.org, oralhealthplatform.eu, pae-eu.eu, siope.eu, smokefreepartnership.eu, uicc.org, vaccineseurope.eu, wemos.nl, zigarettenverband.de
Social Economic Interests.-76	adm-ev.de, dianova.pt, ensie.org, eu.org, gsef-net.org, hitachi.eu, kbs-frb.be, nesst.org, nesta.org.uk, socialfinance.org.uk, veblen-institute.org

Transportation industry advocacy.-77	abta.com, acem.eu, airlines.org, airportaar.ro, anabac.org, aopa.de, asaworld.aero, bovag.nl, cer.be, clepa.eu, ebaa.org, ebma-brussels.eu, ectaa.org, etsc.eu, europeanshippers.eu, evofenedex.nl, gama.aero, gbta.org, gracq.org, iata.org, interferry.com, iru.org, konfederacjalewiatan.pl, passengerrightsadvocates.eu, pfa-auto.fr, posteurop.org, raildeliverygroup.com, rederi.no, seaeurope.eu, sme4space.org, smmt.co.uk, tlp.org.pl, transportforetagen.se, uetr.eu, unife.org, vdr-service.de, vdv.de
Competitiveness interest groups.-78	cecimo.eu, cmc-cvc.com, danskehavne.dk, employers.ee, ert.eu, fairsearch.org, gradiant.org, pih.org.pl, sete.gr
Miscellaneous Technology and Education - 79	ae-info.org, apre.it, claire-ai.org, efa-aef.eu, etc-corporate.org, eucor-uni.org, feam.eu, garagerasmus.org
Youth advocacy and representation.-80	acpypn.com, bjr.de, bjv.at, dbjr.de, europeanhorizons.org, fyeg.org, gceurope.org, yeenet.eu, youngdemocrats.eu, youthcancereurope.org, youthepp.eu
Interest groups related to energy/electronics.-81	aelec.es, aeneas-office.org, amdea.org.uk, amps.org.uk, batteryinnovation.org, beama.org.uk, chademo.com, chargeupeurope.eu, cobaltinstitute.org, currenteurope.eu, eera-set.eu, ehpa.org, ember-climate.org, epbaeurope.net, epeeglobal.org, eurelectric.org, eurima.org, eurobat.org, eurovent.eu, feta.co.uk, iald.org, lightingeurope.org, meca.org, nickelinstitute.org, pvthin.org, rechargebatteries.org, solarheateurope.eu
Banking Industry interest groups.-82	abbl.lu, almega.se, anfia.it, bankenverband.de, bdli.de, bitkom.org, cepli.eu, charteredaccountants.ie, eaca.eu, ecaeurope.com, fcio.at, fiec.eu, fim.net, hypo.org, mittelstandsverbund.de, nvm.nl, uepg.eu, weceurope.org, wir-leben-genossenschaft.de

Financial Industry Interest Groups.-83	acb.com.cy, accaglobal.com, aebanca.es, aipb.it, atmia.com, baft.org, bankwatch.org, bpfi.ie, bpi.com, bsa.org.uk, counter-balance.org, die-dk.de, ebf.eu, fanet.dk, febea.org, fecif.eu, fenca.eu, fia.org, fla.org.uk, foreignbanks.org.uk, gsfc-germany.com, hub.hr, ieaf.es, mfc.org.pl, millenniumbcp.pt, paymentseurope.eu, positivemoney.eu, schuldenberatung.at, swedishbankers.se, swissfinancecouncil.org, vab.de, wsbi-esbg.org
International cultural education and advocacy.-84	annalindhfoundation.org, britishcouncil.org, culture-actioneurope.org, encatc.org, eunicglobal.eu, fingo.fi, ibo.org, moreeurope.org, universitiesuk.ac.uk, yfu.org, ypfp.org
Consumer advocacy groups.-85	aim.be, area-eur.be, betterfinance.eu, beuc.eu, brc.org.uk, britishbrandsgroup.org.uk, cecu.es, consumerchoicecenter.org, consumersforum.it, ekpizo.gr, eurocoop.coop, foodwatch.org, iacc.org, nrf.com, sospotrebitelov.sk, sverigeskonsumenter.se, tacd.org, theconsumergoodsforum.com, verbraucherzentrale.nrw
Intellectual property advocacy.-86	aippi.org, apram.com, bapla.org.uk, britishcopyright.org, ecta.org, fair-standards.org, ficpi.org, inta.org, internationalpublishers.org, ip2innovate.eu, ipo.org, leistungsschutzrecht.info
Media-related interest groups.-87	acte.be, cmfe.eu, confindustriaradiotv.it, disinfo.eu, egta.com, europa-distribution.org, freepressunlimited.org, ifta-online.org, ipi.media, magazinemedia.eu, mertek.eu, newsmediaeurope.eu, sne.fr, the-aop.org, unic-cinemas.org, vlaamsenieuwsmedia.be
Advocacy for Openness and Free Knowledge.-88	april.org, aspeninstitute.de, eclipse.org, fsfe.org, oaspa.org, opensocietyfoundations.org, rd-alliance.org, webfoundation.org
Technology and Innovation.-89	aioti.eu, ametic.es, ectp.org, elra.info, enoll.org, etn.global, etp4hpc.eu, eu-robotics.net, eubac.org, ieee.org, medtecheurope.org, nereus-regions.eu, nlaic.com, pole-optitec.com, smartfactory.de, spectaris.de, techworks.org.uk, vplt.org, xbrleurope.org, zpav.pl, zvo.org

Interest groups fighting poverty.- 90	armutskonferenz.at, atd-quartmonde.org, care-international.org, care.de, caritas.eu, charitytax-group.org.uk, cordaid.org, danchurchaid.org, eapn.eu, eapn.ie, habitat.org, ri.org, snv.org, welthunger-hilfe.de
Miscellaneous European Industries - 91	aegiseurope.eu, culturalfoundation.eu, dkgev.de, ec-spm.org, espo.be, esu-online.org, etoa.org, eu-roalter.com, european-net.org, europeanlawinsti-tute.eu, europeantheatre.eu, franceindustrie.org, jef.eu, mouvement-europeen.eu, ne-mo.org, netzwerk-ebd.de, obessu.org, ordoiuris.pl, uroweb.org, we-move.eu
EU business relations worldwide - 92	aebrus.ru, ccceu.eu, eba.am, elnetwork.eu, eufoa.org, euroamerica.org, eurocham-cambodia.org, eurocham-myanmar.org, eurocham.org.sg, eurochile.cl, eu-romediter.eu, europeindia.eu
Interest groups related to fishing.-93	bluefisheurope.org, cepesca.es, cffacape.org, fishsec.org, ipnlf.org, sff.co.uk, sjomatnorge.no, swfpa.com, visned.nl
Advocacy for policy reform.-94	bepsmonitoringgroup.org, citywide.ie, diogenis.info, encod.org, euro-yoda.org, eurovia.org, ibanet.org, ik.org.pl, penalreform.org, pewtrusts.org
Diverse sustainability interest groups.-95	alianzaporlasolidaridad.org, bellona.org, cepi.org, dgnb.de, dn.dk, eaireland.com, ebcd.org, econsense.de, eeac.eu, eiif.org, estep.eu, eurococoa.com, fer-ver.eu, global2000.at, globalreporting.org, greenovate-europe.eu, ibu-epd.com, idhsustainabletrade.com, iidma.org, keepnorthernirelandbeautiful.org, kidv.nl, letsdoitfoundation.org, mio-ecsde.org, mtvsz.hu, natu-urenmilieu.nl, necstour.eu, nf-int.org, nvc.nl, phospho-rusplatform.eu, power-shift.de, rainforest-alliance.org, sasb.org, somo.nl, startupprize.eu, swedisol.se, thecon-creteinitiative.eu, value-balancing.com, worldbench-markingalliance.org, worldgbc.org, wrforum.org, wwf.eu, wwf.gr

Climate interest groups.-96	carbonmarketwatch.org, cdsb.net, changepartnership.org, climatebonds.net, climatestrategies.org, corporateleadersgroup.com, ieta.org, iigcc.org, klima-allianz.de, mwv.de, negative-emissions.org, regions20.org, reseauactionclimat.org, sandbag.be, stiftung2grad.de, theclimategroup.org
Economic Development Groups.-97	amcham.ro, businesseurope.eu, businessmed-umce.org, edfi.eu, fondromania.org, icaafrica.coop, ihk-muenchen.de, insuleur.org, iticnet.org, ktto.net, linpra.lt, sbe.org.gr, sloga-platform.org
Youth Empowerment Advocates.-98	aegee.org, ecyc.org, eryica.org, eurodesk.eu, iglyo.com, issa.nl, uniarozwoju.org.pl, wagggs.org, youthforum.org
Financial advocacy groups.-99	afme.eu, amafi.fr, aref.org.uk, assogestioni.it, bvai.de, dufas.nl, eemua.org, european-microfinance.org, eurosif.org, fdata.global, finance-watch.org, finansnorge.no, fondbolagen.se, gfma.org, giia.net, investuotojams.eu, pensionseurope.eu, plsa.co.uk, shareaction.org, sifma.org, theaic.co.uk, thecityuk.com, ukfinance.org.uk, voeig.at

Table B.1: All Lobby Clusters with Member Domains

C Law-Making

In the following sections, we give the top-50 words and bigrams predictive of acceptance and rejection of a law edit (c.f. Chapter 5).

Terms Predictive of Acceptance

Words Added

berec | fishing | should | equipment | office | registered | advisory | inserted | important | actions | 2018 | bargaining | best | therefore | transparency | regulators | fisheries | positive | withdrawal | plan | x | gender | financial | ppe | lisa | communication | defence | ” | fuel | second | external | toll | processes | common | buyer | skills | inform | reduce | digital | impact | 2005 | pension | v | contributions | support | council | fitting | agricultural | investigation | processing

Words Removed

berec | safety | eurojust | breeding | surveillance | area | council | consumers | human | 2 | authorised | powers | bodies | hosting | animals | articles | conditions | derogation | ; | 29 | medium | if | manufacturer | origin | allocated | | audit | implementation | provision | conformity | added | specific | 10 | plant | fitting | representative | action | fisheries | amending | 8 | current | financing | political | chapter | identifying | virtual | during | harm | compensation | breed

Context Words

” | appliance | appliances | controls | rco | prima | alternative | harmonised | threats | voice | egf | safety | eurojust | iccat | manufacturer | 63 | breeding | published | engines |

customs | instrument | outside | associated | instructions | creditors | fittings | processed | destination | audit | uniform | number | cash | operating | notified | recipients | positive | institutions | appeal | alcohol | observations | berec | multi | _____ | guidelines | accounts | practical | expenditure | firms | ecosystems | saving

Title Words

community | customs | DDDD-DDDD | control | mediterranean | service | 'customs | supervision | installations | parliament | recovery | cableway | pollutants | multiannual | field | annex | equipment | temporary | council | documents | competition | anti-fraud | statistics | area | drinks | animals | appliances | DDDD/DDD | burning | gaseous | ukraine | fuels | it | policy | agency | zootechnical | plan | laws | ensure | other | fisheries | genealogical | spirit | financial | authorities | DD/DDDD | office | insolvency | investigations | management

Bigrams Added

their sector | opposition , | of meeting | the berec | , humification | this regulation | avoid social | this expert | way behind | a . | berec office | transmission of | in easy | were neither | . 2 | applicable the | violence is | eu´s greenhouse | one of | risk premia | where applicable | economic operators | positive impact | within the | multinationals at | relative deviation | institution , | is inserted | by sub | not properly | accept , | regions , | further amended | the third | complaint was | or federal | people and | , raising | family associations | intelligent mobility | - carrier | carrier economic | acoustic signals | , 51 | elected a | inserted : | signal processing | board of | 2018 prices | transport agency

Bigrams Removed

2 . | . 2 | . . | international efforts | , member | . where | the following | 5 . | hosting service | the case | ; the | human rights | the member | market surveillance | requirements of | data protection | subject to | in such | to that | 4 . | : the | the hosting | to be | which are | plan ; | evaluations ; | whether the | of that | covered by | and related | sharing and | the council | relevant for | eurojust shall | conditions , | . 3 | service provider | the development | provided for | the implementing | in hormonal | notified as | the efsd | that the | article 11 | any other | 6 . | of an | deleted . | take a

Context Bigrams

. ' | : ' | . those | 2 , | . 3 | . 2 | notified body | . " | under other | authority to | requirements of | annual work | their citizenship | ' the | management board | economic

operators | is in | renewable energy | public sector | year . | 3 . | supervisory authorities | investment firms | quantified , | be deferred | the egf | and shall | regulation . | voice communications | this paragraph | promoter . | within a | , storage | / 22 | in other | shall : | of participants | authorities should | monitor the | of new | states remain | as set | ' interests | financing types | year , | article 38 | commission in | consumers ' | resident or | the institutions

Title Bigrams

council on | to regulation | cableway installations | supervision of | ' programme | multiannual recovery | 'customs ' | of customs | , (| european parliament | the 'customs | recovery plan | and of | annex a | field of | parliament and | the reform | on insolvency | insolvency proceedings | a to | replacing annex | the field | and establishing | for cooperation | general budget | control equipment | customs control | budget of | rules applicable | and administrative | gaseous fuels | the mediterranean | DDDD/DDD on | financial rules | the use | in and | for trade | burning gaseous | procedures for | regulation of | the council | and supervision | to the | plan for | appliances burning | no DD/DDDD | zootechnical and | spirit drinks | medicines agency | european medicines

Terms Predictive of Rejection

Words Added

cabotage | these | deleted | ; | eu | societal | must | mercury | payment | illegal | benchmark | territorial | e | hydrogen | except | asylum | – | commercial | service | according | operational | include | basic | agreements |) | additionality | determined | case | consent | circumstances | after | ten | days | constant | negative | firearms | above | s | if | professionals | children | surveillance | set | only | settlement | amended | medical | hours | defined

Words Removed

energy | should | migration | additional | competitiveness | public | workers | corps | irregular | % | different | product | systems | international | forest | eib | efsi | remuneration | growth | joint | before | research | worker | economic | electronic | therefore | passenger | matter | works | solidarity | months | value | reporting | through | provide | online | can | eurodac | impact | monitoring | allowances | every | identity | account | cultural | concerted | supply | projects | format | structural

Context Words

_____ | allocation | rightholders | firearms | hosting | posting | allocations | resettlement | reserve | allowances | benchmark | foreign | core | free | labels | pnr | educational | verification | driver | collective |) | works | investments | 2030 | preservation | terrorist | forest | solidarity | remote | advanced | mercury | 25 | broadcast | ancillary | fingerprints | million | employees | condition | redress | settings | excellence | parental | 5% | penalties | travel | seller | exception | enisa | renovation | containing

Title Words

and | directive | market | DDDD | framework | services | DDDD/DDD/ec | requirements | </s> | protection | agricultural | energy | as | a | for | gas | decision | greenhouse | contracts | digital | operation | online | name | strategic | development | emission | regulation | view | of | from | at | georgia | instruments | establishment | structural | trading | record | plans | in | investments | supplementary | regards | the | specific | private | relevant | copyright | DDD/DDDD | pnr | posting

Bigrams Added

“10a . | communication , | normalisation process | welfare regulations | that activity | is deleted | become apparent | general production | ; | certificates were | and in | may propose | with the | and logistics | ’ s | separation of | engine replacement | different generators | . in | hatred . | according to | or morality | as authors | made explicitly | value cases | , point | annex , | as jointly | lifting a | climate transition | - contributions | parties to | 32a is | valued by | place of | fuels for | directly awarding | a minimum | judges each | procedures overcoming | european union | ii may | new genetic | is amended | healthcare professionals | service + | , interpreters | leave may | between solid | status under

Bigrams Removed

no reason | and now | terrorism - | as the | contribute to | digital content | , possessed | other subject | guidance , | under the | electronic monitoring | the digital | . this | agricultural guarantee | than ten | legal body | which the | and other | , shall | 0 . | remote electronic | the cir | . member | the supplier | the passenger | authorised periods | solidarity corps | . in | in case | the product | ’ association | of directive | policy objectives | the forest | - sharing | discussion . | the amount |] and | they are | intention or | 1 . | least likely | information on | same shall | - matter | state to | and of | carbon impact | union law | ; and

Context Bigrams

contents of | a sub | therefore , | hosting service | . _____ | the funds | of directive | pnr data | 000 for | parental leave | report within | regional operational | commission may | article 4 | investment board | paragraph 3 | produced from | procedure , | scientific evaluation | works or | their common | the driver | main third | or other | free allocation | state which | 27 . | deemed to | ' shall | states introduce | have given | . member | states may | data for | their rights | they shall | finance may | for free | programme's research | commission should | article 2 | - use | may request | down rules | consumers , | authority referred | the supply | which establishes | - and | be deemed

Title Bigrams

) and | and regulation | directive DDDD/DDD/ec | greenhouse gas | corps programme | rules for | european union | services in | , regulation | the eu | passenger name | eu pnr | by member | DDDD/DDD and | the structural | structural reform | name record | of passenger | data (| the framework | efficiency labelling | strategic plans | to georgia | for screening | european agricultural | DDDD/DDDD with | as regards | DDDD/DDD/ec and | record data | council amending | and weekly | DDD/DDDD as | daily and | directive DDDD/DD/eu | , and | a framework | pnr) | internal market | rules on | concerning the | screening of | protection certificate | for medicinal | supplementary protection | in criminal | , laying | against dumped | DDDD/DDDD on | of energy | of a

Bibliography

Adler, B. T., & de Alfaro, L. (2007). A content-driven reputation system for the Wikipedia. *Proceedings of WWW'07*.

Agin, S., & Karlsson, M. (2021). Mapping the field of climate change communication 1993–2018: Geographically biased, theoretically narrow, and methodologically limited. *Environmental Communication*, *15*(4), 431–446.

Allen Institute for AI. (2022). longformer-base-4096 [Accessed: 2022-08-1]. https://huggingface.co/allenai/longformer-base-4096

AllSides. (2022). AllSides.com [Accessed: 2022-09-15]. https://www.allsides.com/unbiased-balanced-news

Arora, S., Liang, Y., & Ma, T. (2016). A simple but tough-to-beat baseline for sentence embeddings. *International Conference on Learning Representations*.

Aroyo, L., Dixon, L., Thain, N., Redfield, O., & Rosen, R. (2019). Crowdsourcing subjective tasks: The case study of understanding toxicity in online discussions. *Companion Proceedings of The 2019 World Wide Web Conference*, 1100–1105. https://doi.org/10.1145/3308560.3317083

Athiwaratkun, B., & Wilson, A. (2017). Multimodal word distributions. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1645–1656.

Athiwaratkun, B., Wilson, A., & Anandkumar, A. (2018). Probabilistic fastText for multi-sense word embeddings. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1–11.

Bahdanau, D., Cho, K. H., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. *3rd International Conference on Learning Representations, ICLR 2015*.

Baller, I. (2017). Specialists, party members, or national representatives: Patterns in co-sponsorship of amendments in the European Parliament. *European Union Politics*, *18*(3), 469–490.

Bartunov, S., Kondrashkin, D., Osokin, A., & Vetrov, D. (2016). Breaking sticks and ambiguities with adaptive skip-gram. *Artificial Intelligence and Statistics*, 130–138.

Bednáriková, Z., & Jílková, J. (2012). Why is the agricultural lobby in the european union member states so effective? *E+M Ekonomie a Management*, (2), 26.

Beltagy, I., Peters, M. E., & Cohan, A. (2020). Longformer: The long-document Transformer. *arXiv preprint arXiv:2004.05150*.

Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, *5*(2), 157–166.

Bennett, W. L. (2016). *News: The politics of illusion*. University of Chicago Press.

Beyer, K., Goldstein, J., Ramakrishnan, R., & Shaft, U. (1999). When is "nearest neighbor" meaningful? *Database Theory—ICDT'99: 7th International Conference Jerusalem, Israel, January 10–12, 1999 Proceedings 7*, 217–235.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, *3*(Jan), 993–1022.

Bloomfield, E. F., & Tillery, D. (2019). The circulation of climate change denial online: rhetorical and networking strategies on facebook. *Environmental Communication*, *13*(1), 23–34.

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

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017b). fastText: Efficient learning of word representations and sentence classification. https://github.com/facebookresearch/fastText

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017c). Wiki word vectors. https://fasttext.cc/docs/en/pretrained-vectors.html

Boldyreva, E. L., Grishina, N. Y., Duisembina, Y., L Boldyreva, E., Y Grishina, N., et al. (2018). Cambridge Analytica: Ethics and online manipulation with decision-making process. *European Proceedings of Social and Behavioural Sciences*, *51*.

Bouwen, P. (2003). A theoretical and empirical study of corporate lobbying in the European Parliament. *European integration online papers (EIoP)*, *7*(11).

Bowman, S. R., Angeli, G., Potts, C., & Manning, C. D. (2015). A large annotated corpus for learning natural language inference. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 632–642. https://doi.org/10.18653/v1/D15-1075

Bradley, R. A., & Terry, M. E. (1952). Rank analysis of incomplete block designs: i. the method of paired comparisons. *Biometrika*, *39*(3/4), 324–345. Retrieved March 29, 2023, from http://www.jstor.org/stable/2334029

Bražinskas, A., Havrylov, S., & Titov, I. (2018). Embedding words as distributions with a Bayesian skip-gram model. *Proceedings of the 27th International Conference on Computational Linguistics*, 1775–1789. https://aclanthology.org/C18-1151

Byrd, R. H., Lu, P., Nocedal, J., & Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, *16*(5), 1190–1208.

Cer, D., Yang, Y., Kong, S.-y., Hua, N., Limtiaco, N., St. John, R., Constant, N., Guajardo-Cespedes, M., Yuan, S., Tar, C., Strope, B., & Kurzweil, R. (2018).

Universal sentence encoder for English. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 169–174. https://doi.org/10.18653/v1/D18-2029

Chalkidis, I., & Kampas, D. (2019). Deep learning in law: Early adaptation and legal word embeddings trained on large corpora. *Artificial Intelligence and Law*, *27*(2), 171–198.

Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J., & Blei, D. (2009). Reading tea leaves: How humans interpret topic models. *Advances in Neural Information Processing Systems*, *22*.

Chen, W.-F., Wachsmuth, H., Al-Khatib, K., & Stein, B. (2018). Learning to flip the bias of news headlines. In A. Gatt, M. Goudbeek, & E. Krahmer (Eds.), *11th international natural language generation conference (inlg 2018)* (pp. 79–88). Association for Computational Linguistics. http://aclweb.org/anthology/W18-6509

Cho, K., van Merriënboer, B., Bahdanau, D., & Bengio, Y. (2014). On the properties of neural machine translation: encoder-decoder approaches. *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*, 103–111.

Chomsky, N. (1965). Aspects of the theory of syntax (vol. 11). *MIT Press. doi*, *10*, 90008–5.

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *NIPS 2014 Workshop on Deep Learning, December 2014*.

Coman, E. E. (2009). Reassessing the influence of party groups on individual members of the European Parliament. *West European Politics*, *32*(6), 1099–1117.

Conneau, A., Kiela, D., Schwenk, H., Barrault, L., & Bordes, A. (2017). Supervised learning of universal sentence representations from natural language inference data. *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 670–680. https://doi.org/10.18653/v1/D17-1070

Costello, R., & Thomson, R. (2010). The policy impact of leadership in committees: Rapporteurs' influence on the European Parliament's opinions. *European Union Politics*, *11*(2), 219–240.

Dahal, B., Kumar, S. A., & Li, Z. (2019). Topic modeling and sentiment analysis of global climate change tweets. *Social network analysis and mining*, *9*, 1–20.

De Kock, C., & Vlachos, A. (2022). Leveraging Wikipedia article evolution for promotional tone detection. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 5601–5613. https://doi.org/10.18653/v1/2022.acl-long.384

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by Latent Semantic Analysis. *Journal of the American Society for Information Science*, *41*(6), 391–407.

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

Dewenter, R., Linder, M., & Thomas, T. (2019). Can media drive the electorate? The impact of media coverage on voting intentions. *European Journal of Political Economy*, *58*, 245–261.

Dialer, D., & Richter, M. (2019). Lobbying in Europe: Professionals, politicians, and institutions under general suspicion? *Lobbying in the European Union: Strategies, Dynamics and Trends*, 1–18.

Dieng, A. B., Ruiz, F. J., & Blei, D. (2020). Topic modeling in embedding spaces. *Transactions of the Association for Computational Linguistics*, *8*, 439–453.

Druck, G., Miklau, G., & McCallum, A. (2008). Learning to predict the quality of contributions to Wikipedia. *Proceedings of WikiAI 2008*.

Duchi, J., Hazan, E., & Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, *12*(Jul), 2121–2159.

Elmimouni, H., Forte, A., & Morgan, J. (2022). Why people trust Wikipedia articles: Credibility assessment strategies used by readers. *Proceedings of the 18th International Symposium on Open Collaboration*. https://doi.org/10.1145/3555051.3555052

Elo, A. E. (1978). *The rating of chess players, past and present*. Arco Pub.

Etter, V., Herzen, J., Grossglauser, M., & Thiran, P. (2014). Mining democracy. *Proceedings of the second ACM conference on Online social networks*, 1–12.

European Parliament. (2019). Ep approves more transparency and efficiency in its internal rules [Accessed: 2023-08-06]. https://www.europarl.europa.eu/news/en/press-room/20190123IPR24128/ep-approves-more-transparency-and-efficiency-in-its-internal-rules

European Parliament. (2021a). European Parliament Portal. https://www.europarl.europa.eu/portal/en

European Parliament. (2021b). Rules of procedure of the European Parliament - Rule 180 [Accessed: 2021-02-14]. https://www.europarl.europa.eu/doceo/document/RULES-9-2019-07-02-RULE-180_EN.html

European Union. (2011). EU Transparency Register [Accessed: 2023-06-20]. https://ec.europa.eu/transparencyregister/public/homePage.do

European Union. (2021). European Data Portal [Accessed: 2021-02-14]. https://www.europeandataportal.eu/en

Firth, J. (1957). A synopsis of linguistic theory, 1930-1955. *Studies in Linguistic Analysis*, 10–32.

Flynn, C., Yamasumi, E., Fisher, S., Snow, D., Grant, Z., Kirby, M., Browning, P., Rommerskirchen, M., & Russell, I. (2021). *Peoples' Climate Vote: Results*. United Nations Development Programme.

Google. (2013). word2vec [Accessed: 2020-10-19]. https://code.google.com/archive/p/word2vec/

Greenstein, S., & Zhu, F. (2012). *Collective intelligence and neutral point of view: the case of Wikipedia* (tech. rep.). National Bureau of Economic Research.

Grootendorst, M. (2022). BERTopic: neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*.

Gustafson, A., Ballew, M. T., Goldberg, M. H., Cutler, M. J., Rosenthal, S. A., & Leiserowitz, A. (2020). Personal stories can shift climate change beliefs and risk perceptions: The mediating role of emotion. *Communication Reports*, *33*(3), 121–135.

Guu, K., Hashimoto, T. B., Oren, Y., & Liang, P. (2018). Generating sentences by editing prototypes. *Transactions of the Association for Computational Linguistics*, *6*, 437–450.

Halfaker, A., & Geiger, R. S. (2020). ORES: Lowering barriers with participatory machine learning in Wikipedia. *Proceedings of the ACM on Human-Computer Interaction*, *4*(CSCW2), 1–37.

Harari, Y. N. (2015). *Sapiens: A brief history of humankind*. Harper Perennial.

Harris, Z. S. (1954). Distributional structure. *WORD*, *10*(2-3), 146–162. https://doi.org/10.1080/00437956.1954.11659520

Hix, S. (2002). Parliamentary behavior with two principals: Preferences, parties, and voting in the European Parliament. *American Journal of Political Science*, 688–698.

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780.

Hofmann, T. (1999). Probabilistic Latent Semantic Indexing. *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 50–57.

HTCondor. (2023). Htcondor overview [Accessed: 2023-08-06]. https://htcondor.org/htcondor/overview/

Ibenskas, R., & Bunea, A. (2021). Legislators, organizations and ties: Understanding interest group recognition in the European Parliament. *European Journal of Political Research*, *60*(3), 560–582.

Immer, A., Kristof, V., Grossglauser, M., & Thiran, P. (2020). Sub-matrix factorization for real-time vote prediction. *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2280–2290.

Integrity Watch. (2023). Integrity Watch Data Hub [Accessed: 2023-06-20]. https://data.integritywatch.eu/

Jiang, Y., Adams, B., & German, D. M. (2013). Will my patch make it? And how fast? Case study on the Linux kernel. *Proceedings of MSR 2013*.

Joachims, T. (1998). Text categorization with support vector machines: Learning with many relevant features. *European Conference on Machine Learning*, 137–142.

Jolly, S., Bakker, R., Hooghe, L., Marks, G., Polk, J., Rovny, J., Steenbergen, M., & Vachudova, M. A. (2022). Chapel Hill Expert Survey trend file, 1999–2019. *Electoral Studies*, *75*, 102420. https://doi.org/https://doi.org/10.1016/j.electstud.2021.102420

Jordan, M. I., Ghahramani, Z., Jaakkola, T. S., & Saul, L. K. (1999). An introduction to variational methods for graphical models. *Machine Learning*, *37*, 183–233.

Joulin, A., Grave, E., Bojanowski, P., Douze, M., Jégou, H., & Mikolov, T. (2016). fastText.zip: Compressing text classification models. *arXiv preprint arXiv:1612.03651*.

Joulin, A., Grave, E., Bojanowski, P., & Mikolov, T. (2016). fastText: Language identification. https://fasttext.cc/docs/en/language-identification.html

Joulin, A., Grave, É., Bojanowski, P., & Mikolov, T. (2017). Bag of tricks for efficient text classification. *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, 427–431.

Keegan, B., & Fiesler, C. (2017). The evolution and consequences of peer producing Wikipedia's rules. *Proceedings of the International AAAI Conference on Web and Social Media*, *11*(1), 112–121.

Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer*, *42*(8), 30–37.

Kreppel, A. (1999). What affects the European Parliament's legislative influence? An analysis of the success of EP amendments. *JCMS: Journal of Common Market Studies*, *37*(3), 521–537.

Kreppel, A. (2002). Moving beyond procedure: An empirical analysis of European Parliament legislative influence. *Comparative Political Studies*, *35*(7), 784–813.

Kristof, V. (2021a). *Discrete-choice mining of social processes* (Doctoral dissertation). EPFL.

Kristof, V. (2021b). Predikon. https://www.predikon.ch/en/

Kristof, V., Grossglauser, M., & Thiran, P. (2020). War of words: The competitive dynamics of legislative processes. *Proceedings of The Web Conference 2020*, 2803–2809.

Kristof, V., Suresh, A., Grossglauser, M., & Thiran, P. (2021). War of words II: Enriched models of law-making processes. *Proceedings of the Web Conference 2021*, 2014–2024. https://doi.org/10.1145/3442381.3450131

Kurtovic, B. (2022). mwparserfromhell [Accessed: 2022-10-14]. https://github.com/earwig/mwparserfromhell

Lefkofridi, Z., & Katsanidou, A. (2014). Multilevel representation in the European Parliament. *European Union Politics*, *15*(1), 108–131.

Leiserowitz, A., Carman, J., Buttermore, N., Neyens, L., Rosenthal, S., Marlon, J., Schneider, J., & Mulcahy, K. (2022). *International public opinion on climate change 2022*. YPCCC: Yale Program on Climate Change Communication.

Li, J., & Jurafsky, D. (2015). Do multi-sense embeddings improve natural language understanding? *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 1722–1732.

Li, Z., Lu, Z., & Yin, M. (2022). Towards better detection of biased language with scarce, noisy, and biased annotations. *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, 411–423.

Lim, S., Jatowt, A., Färber, M., & Yoshikawa, M. (2020). Annotating and analyzing biased sentences in news articles using crowdsourcing. *Proceedings of the Twelfth Language Resources and Evaluation Conference*, 1478–1484. https://aclanthology.org/2020.lrec-1.184

Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis*. Wiley.

Maaten, L. v. d., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, *9*(Nov), 2579–2605.

Matei, S. A., & Dobrescu, C. (2011). Wikipedia's "neutral point of view": Settling conflict through ambiguity. *The Information Society*, *27*(1), 40–51.

Maystre, L., Kristof, V., & Grossglauser, M. (2019). Pairwise comparisons with flexible time-dynamics. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 1236–1246. https://doi.org/10.1145/3292500.3330831

McInnes, L., Healy, J., Saul, N., & Großberger, L. (2018). UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software*, *3*(29), 861.

Meng, Y., Zhang, Y., Huang, J., Zhang, Y., & Han, J. (2022). Topic discovery via latent space clustering of pretrained language model representations. *Proceedings of the ACM Web Conference 2022*, 3143–3152.

Metaxas, P., Mustafaraj, E., Wong, K., Zeng, L., O'Keefe, M., & Finn, S. (2015). What do retweets indicate? Results from user survey and meta-review of research. *Proceedings of the International AAAI Conference on Web and Social Media*, *9*(1), 658–661.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, *26*.

Mimno, D., Wallach, H., Talley, E., Leenders, M., & McCallum, A. (2011). Optimizing semantic coherence in topic models. *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, 262–272. https://aclanthology.org/D11-1024

Mnih, A., & Teh, Y. W. (2012). A fast and simple algorithm for training neural probabilistic language models. *Proceedings of the 29th International Conference on Machine Learning*, 419–426.

Moller-Nielsen, T. (2023). Scandal at European Parliament: New leaked file reveals sweeping Qatargate corruption. https://www.brusselstimes.com/608211/scandal-at-european-parliament-new-leaked-file-reveals-sweeping-qatargate-corruption

Mu, J., Bhat, S., & Viswanath, P. (2017). Representing sentences as low-rank subspaces. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 629–634. https://doi.org/10.18653/v1/P17-2099

Mühlböck, M. (2012). National versus European: Party control over members of the European Parliament. *West European Politics*, *35*(3), 607–631.

Neelakantan, A., Shankar, J., Passos, A., & McCallum, A. (2014). Efficient non-parametric estimation of multiple embeddings per word in vector space. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1059–1069.

Obama White House. (2018). Open Government Initiative [Accessed: 2020-10-19]. https://obamawhitehouse.archives.gov/open

OpenAI. (2022). ChatGPT [Accessed: 2023-07-10]. https://chat.openai.com/

OpenAI. (2023a). Chat Completions API [Accessed: 2023-06-20]. https://platform.openai.com/docs/guides/gpt/chat-completions-api

OpenAI. (2023b). GPT-4 technical report.

Parltrack. (2023). Parltrack. https://parltrack.org/

Pavalanathan, U., Han, X., & Eisenstein, J. (2018). Mind your POV: Convergence of articles and editors towards Wikipedia's neutrality norm. *Proceedings of the ACM on Human-Computer Interaction*, *2*(CSCW), 1–23.

Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized word representations. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, 2227–2237. https://doi.org/10.18653/v1/N18-1202

Pew Research. (2011). Press widely criticized, but trusted more than other information sources. https://www.pewresearch.org/politics/2011/09/22/press-widely-criticized-but-trusted-more-than-other-institutions/

Pryzant, R., Diehl Martinez, R., Dass, N., Kurohashi, S., Jurafsky, D., & Yang, D. (2020). Automatically neutralizing subjective bias in text. *Proceedings of the AAAI Conference on Artificial Intelligence*, *34*(01), 480–489. https://doi.org/10.1609/aaai.v34i01.5385

Qi, Y., Sachan, D., Felix, M., Padmanabhan, S., & Neubig, G. (2018). When and why are pre-trained word embeddings useful for neural machine translation? *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, 529–535.

Qian, E. (2019). Twitter Climate Change Sentiment Dataset [Accessed: 2022-08-1]. https://www.kaggle.com/datasets/edqian/twitter-climate-change-sentiment-dataset

Rasmussen, M. K. (2015). The battle for influence: The politics of business lobbying in the European Parliament. *JCMS: Journal of Common Market Studies*, *53*(2), 365–382.

Reimers, N. (2022). EasyNMT. https://github.com/UKPLab/EasyNMT

Reimers, N., & Gurevych, I. (2019a). Sentence-BERT: Sentence embeddings using siamese BERT-networks. *Proceedings of the 2019 Conference on Empirical Methods in*

Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 3982–3992.

Reimers, N., & Gurevych, I. (2019b). Sentence-Transformers. https://www.sbert.net/index.html

Sarkar, S., Reddy, B. P., Sikdar, S., & Mukherjee, A. (2019). StRE: Self attentive edit quality prediction in Wikipedia. *Proceedings of ACL 2019*, 3962–3972.

Schmidhuber, J. (1992). Learning complex, extended sequences using the principle of history compression. *Neural Computation*, *4*(2), 234–242. https://doi.org/10.1162/neco.1992.4.2.234

Sia, S., Dalmia, A., & Mielke, S. J. (2020). Tired of topic models? clusters of pretrained word embeddings make for fast and good topics too! *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1728–1736. https://doi.org/10.18653/v1/2020.emnlp-main.135

Solaiman, B. (2023). Lobbying in the uk: towards robust regulation. *Parliamentary Affairs*, *76*(2), 270–297.

Spinde, T., Rudnitckaia, L., Mitrović, J., Hamborg, F., Granitzer, M., Gipp, B., & Donnay, K. (2021). Automated identification of bias inducing words in news articles using linguistic and context-oriented features. *Information Processing & Management*, *58*(3), 102505. https://doi.org/https://doi.org/10.1016/j.ipm.2021.102505

Stanford Internet Observatory. (2021). Inauthentic editing: Changing Wikipedia to win elections and influence people. https://cyber.fsi.stanford.edu/io/news/wikipedia-part-one

Sufi, F. K., Razzak, I., & Khalil, I. (2022). Tracking anti-vax social movement using ai-based social media monitoring. *IEEE Transactions on Technology and Society*, *3*(4), 290–299.

Suresh, A., Milikic, L., Murray, F., Zhu, Y., & Grossglauser, M. (2023). Mining effective strategies for climate change communication. *ICLR 2023 Workshop on Tackling Climate Change with Machine Learning*. https://www.climatechange.ai/papers/iclr2023/38

Suresh, A., Radojevic, L., Salvi, F., Magron, A., Kristof, V., & Grossglauser, M. (2023). Studying lobby influence in the European Parliament. *Under review at EMNLP 2023*.

Suresh, A., Wu, C.-H., & Grossglauser, M. (2023). It's all relative: Interpretable models for scoring bias in documents. *arXiv preprint arXiv:2307.08139*.

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. *Advances in neural information processing systems*, *27*.

Swiss Government. (2021). Swiss Open Government Data [Accessed: 2021-02-14]. https://opendata.swiss/en/

Tan, C., Lee, L., & Pang, B. (2014). The effect of wording on message propagation: Topic-and author-controlled natural experiments on Twitter. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 175–185.

Tarrant, A., & Cowen, T. (2022). Big Tech lobbying in the EU. *The Political Quarterly*, *93*(2), 218–226.

Taylor, W. L. (1953). "cloze procedure": a new tool for measuring readability. *Journalism Quarterly*, *30*(4), 415–433.

Tian, F., Dai, H., Bian, J., Gao, B., Zhang, R., Chen, E., & Liu, T.-Y. (2014). A probabilistic model for learning multi-prototype word embeddings. *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, 151–160.

Tiedemann, J., & Thottingal, S. (2020). OPUS-MT — Building open translation services for the world. *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation (EAMT)*.

Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., et al. (2023). LLaMA: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*.

Transparency International. (1993). Mission, Vision and Values [Accessed: 2023-06-20]. https://www.transparency.org/en/the-organisation/mission-vision-values

Tsebelis, G., Jensen, C. B., Kalandrakis, A., & Kreppel, A. (2001). Legislative procedures in the European Union: An empirical analysis. *British Journal of Political Science*, *31*(4), 573–599.

Twitter. (2023). Twitter API. https://developer.twitter.com/en/docs/twitter-api

UN Global Pulse. (2014). Taxonomy for studying climate change tweets [Accessed: 2022-11-4]. http://unglobalpulse.net/climate/

Vafa, K., Naidu, S., & Blei, D. (2020). Text-Based Ideal Points. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5345–5357. https://doi.org/10.18653/v1/2020.acl-main.475

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

Vilnis, L., & McCallum, A. (2014). Word representations via Gaussian embedding. *arXiv preprint arXiv:1412.6623*.

Vu, H. T., Blomberg, M., Seo, H., Liu, Y., Shayesteh, F., & Do, H. V. (2021). Social media and environmental activism: Framing climate change on Facebook by global NGOs. *Science Communication*, *43*(1), 91–115.

Wang, A., & Cho, K. (2019). BERT has a mouth, and it must speak: BERT as a Markov random field language model. *NAACL HLT 2019*, 30.

Wang, K., Bansal, M., & Frahm, J.-M. (2018). Retweet wars: Tweet popularity prediction via dynamic multimodal regression. *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 1842–1851.

Wikimedia. (2023). MediaWiki Action API [Accessed: 2023-07-30]. https://www.mediawiki.org/wiki/API:Main_page

Wikipedia. (2022a). Wikipedia: List of controversial issues [Accessed: 2022-08-01]. https://en.wikipedia.org/wiki/Wikipedia:List_of_controversial_issues

Wikipedia. (2022b). Wikipedia:Manual of StyleWords to Watch [Accessed: 2022-08-01]. https://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style/Words_to_watch

Wikipedia. (2022c). Wikipedia:NPOV [Accessed: 2022-10-14]. https://en.wikipedia.org/wiki/Wikipedia:Neutral_point_of_view

Wikipedia. (2023a). ORES [Accessed: 2023-05-15]. https://www.mediawiki.org/wiki/ORES

Wikipedia. (2023b). Wikipedia: Good article criteria [Accessed: 2023-05-15]. https://en.wikipedia.org/wiki/Wikipedia:Good_article_criteria

Williams, A., Nangia, N., & Bowman, S. (2018). A broad-coverage challenge corpus for sentence understanding through inference. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, 1112–1122. https://doi.org/10.18653/v1/N18-1101

Wong, K., Redi, M., & Saez-Trumper, D. (2021). Wiki-Reliability: A large scale dataset for content reliability on Wikipedia. *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2437–2442.

Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., et al. (2016). Google's neural machine translation system: bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144*.

Xia, Y., Chen, T. H. Y., & Kivelä, M. (2021). Spread of tweets in climate discussions: A case study of the 2019 Nobel Peace Prize announcement. *Nordic Journal of Media Studies*, *3*(1), 96–117.

Yardim, A. B., Kristof, V., Maystre, L., & Grossglauser, M. (2018). Can who-edits-what predict edit survival? *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2604–2613. https://doi.org/10.1145/3219819.3219979

Yin, P., Neubig, G., Allamanis, M., Brockschmidt, M., & Gaunt, A. L. (2018). Learning to represent edits. *International Conference on Learning Representations*.

Zermelo, E. (1929). Die berechnung der turnier-ergebnisse als ein maximumproblem der wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, *30*, 436–460.

Zhang, X., Zhao, J., & LeCun, Y. (2015). Character-level convolutional networks for text classification. *Advances in Neural Information Processing Systems*, 649–657.

Zhang, Z., Fang, M., Chen, L., & Rad, M. R. N. (2022). Is neural topic modelling better than clustering? An empirical study on clustering with contextual embeddings for topics. *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 3886–3893.

Zhong, Y., Yang, J., Xu, W., & Yang, D. (2021). WIKIBIAS: Detecting multi-span subjective biases in language. *Findings of the Association for Computational Linguistics: EMNLP 2021*, 1799–1814.

ASWIN SURESH

Ecublens, Switzerland | Ph: +41-766647967 | aswin.suresh@epfl.ch

EDUCATION

EPFL — Lausanne, Switzerland
Doctor of Philosophy in Computer and Communication Sciences — 2017-2023 (Expected)
Thesis Title: "Interpretable Text-Based Models for Understanding Social Phenomena"
Cumulative GPA: 5.75/6.00
Relevant Coursework: Machine Learning, Deep Learning for NLP, Performance Evaluation

EPFL — Lausanne, Switzerland
Master of Science in Communication Systems — 2014-2017
Thesis Title: "Proactive Fault Recovery for Real-Time Mission-Critical Systems"
Cumulative GPA: 5.35/6.00
Relevant Coursework: Distributed Information Systems, Distributed Algorithms, TCP/IP Networking

IIT JODHPUR — Jodhpur, India
Bachelor of Technology in Electrical Engineering — 2010-2014
Project Title: "Design and Implementation of Global Navigation Satellite System Receiver"
Cumulative GPA: 9.96/10.00; Top ranked in graduating class across all disciplines
Relevant Coursework: Pattern Recognition, Complex Networks, Wireless Communication

INTERNSHIPS

CISCO SYSTEMS — Ecublens, Switzerland
Research Intern (Master) — Feb 2016 – Aug 2016
- Designed a novel and efficient algorithm to perform load balancing of video clients on enterprise wireless networks by executing directed roams while minimizing disruption due to roaming
- Evaluated the algorithm using simulations; by roaming just 5% of clients the algorithm could increase throughput by at least 10% for over half of the clients and by at least 20% for over a quarter of the clients
- Implemented a prototype on a real wireless network
- Presented the work at IEEE WCNC 2018 in Barcelona, Spain

UNIVERSITY OF SOUTHERN CALIFORNIA — Los Angeles, United States
Visiting Scholar (Bachelor), Viterbi School of Engineering — May 2013 – July 2013
- Studied quantum dots and compared their methods of synthesis
- Structurally characterized quantum dots using atomic force microscopy

INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH — Thiruvananthapuram, India
Project Student (Bachelor), School of Physics — May 2012 – July 2012
- Fabricated efficient and reliable organic memory devices with write-once-read-many times (WORM) characteristics
- Published the work in the journal Physical Chemistry Chemical Physics

OTHER PROJECTS

CYBER ATTACK ON TIME SYNCHRONISATION PROTOCOLS — EPFL, Spring Semester 2015
- Set up a test bed synchronizing two desktops with White Rabbit Precision Time Protocol (WR-PTP), a candidate protocol for network synchronization on Smart Grids achieving sub-nanosecond accuracy
- Demonstrated an attack on the synchronization by injecting a one-way delay, resulting in an error of the order of a few microseconds
- Presented the work at IEEE I2MTC 2016 in Taipei, Taiwan

SIMULATION OF HYBRID WIFI/PLC NETWORKS EPFL, Fall Semester 2015

- Simulated WiFi networks on NS3. Studied throughput achieved by different rate control algorithms, under different channel conditions, and with contention
- Modified PLC implementation on NS3 to include Frame Aggregation and Selective Block Acknowledgement

PUBLICATIONS AND OTHER MANUSCRIPTS

- Suresh, A., Radojevic, L., Salvi, F., Magron, A., Kristof, V., & Grossglauser, M. (2023). Studying Lobby Influence in the European Parliament. *Under Review at EMNLP 2023*
- Suresh, A., Wu, C. H., & Grossglauser, M. (2023). It's All Relative: Interpretable Models for Scoring Bias in Documents. *arXiv preprint arXiv:2307.08139*
- Suresh, A., Milikic, L., Murray, F., Zhu, Y., & Grossglauser, M. (2023). Mining Effective Strategies for Climate Change Communication. *ICLR 2023 Workshop on Tackling Climate Change with Machine Learning*
- Kristof, V.*, Suresh, A.*, Grossglauser, M., & Thiran, P. (2021). War of Words II: Enriched Models of Law-Making Processes. *Proceedings of the Web Conference 2021*
- Suresh, A., Mena, S., Tomozei, D. C., Granai, L., Zhu, X., & Ferrari, S. (2018). Load Balancing Video Clients in Enterprise Wireless Networks. *2018 IEEE Wireless Communications and Networking Conference (WCNC)*
- Barreto, S., Suresh, A., & Le Boudec, J. Y. (2016). Cyber-attack on packet-based time synchronization protocols: The undetectable delay box. *2016 IEEE International Instrumentation and Measurement Technology Conference Proceedings*
- Suresh, A., Krishnakumar, G., & Namboothiry, M. A. (2014). Filament theory based WORM memory devices using aluminum/poly (9-vinylcarbazole)/aluminum structures. *Physical Chemistry Chemical Physics*, *16(26)*

TEACHING ASSISTANTSHIPS

- **Internet Analytics (Spring 2019-2022)**
  Prepared exam questions, supervised and graded labs and exams
- **Machine Learning (Fall 2018-2020)**
  Supervised and graded projects, supervised labs and exams
- **Probability and Statistics (Spring 2018)**
  Supervised exercise sessions, graded exams
- **Advanced Digital Communication (Student Assistant) (Fall 2016)**
  Supervised exercise sessions, graded homework

STUDENT PROJECTS SUPERVISED

- **Studying Lobby Influence in the European Parliament**
  Lazar Radojevic and Francesco Salvi (Spring 2023), Pratyush Gupta (Summer 2022), Bayazit Deniz, Bhargav Srinivas, Charlie Castes, Mohammed Allouch, Benedek Harsanyi, Kamil Czerniak (Nov-Dec 2021), Mahmoud Sellami (Spring 2021), Antoine Magron (Fall 2020, Summer 2021)
- **Framing in Wikipedia**
  Chi-Hsuan Wu (Spring, Summer 2022), Weier Liu (Fall 2021)
- **Predicting Swiss Votes through Machine Learning**
  Matthieu Andre (Spring, Summer, Fall 2022), Victor Gergaud (Spring 2022), Thomas Berkane (Fall 2021), Yann Yasser (Spring 2021)
- **Climate Change Framing**
  Francis Murray, Orfeas Liossatos, Henrique Da Silva Gameiro, Lazar Milikic, Marko Lisicic, Yurui Zhu (Nov-Dec 2022), Boran Xu (Summer 2022), Omar El Malki (Spring 2022), Theodoros Bitsakis (Fall 2021)

ADDITIONAL

Awards: EPFL IC Teaching Assistant Award 2019, 2021; EPFL Excellence Fellowship; President's Gold Medal (IIT Jodhpur)
Technical Skills: Python, PyTorch, Tensorflow, Keras, PySpark, Octave, MATLAB, Bash, Git, LaTeX
Languages: English (Fluent); Malayalam (Native); Hindi (Working proficiency); French (A1/A2, Basic)
Hobbies: Playing the violin (Indian classical music), Hiking, Biking