Cover Page – MSc Business Analytics Consultancy Project/Dissertation 2021-22

Title of project: Faces of the Pandemic – An Analysis of Political Sentiment in the Swiss Newspaper Coverage and the Public Discourse on Twitter

Date: 8th of August, 2022

Disclaimer:

I hereby declare that this dissertation is my individual work and to the best of my knowledge and confidence, it has not already been accepted in substance for the award of any other degree and is not concurrently submitted in candidature for any degree. It is the end product of my own independent study except where other acknowledgement has been stated in the text.

Abstract

The Swiss population has shown an extremely critical perception of the Covid-19 coverage in domestic newspapers, with more than 75% believing that they have failed to provide an objective account of the developments during the pandemic [70]. My project seeks to evaluate these charges against the Swiss media in a previously unseen manner. Using natural language processing methods, the political sentiment in newspaper articles and tweets is investigated, allowing for a comparison between traditional and social media as key sources of public opinion. In particular, it is analyzed i) to what extent the mentions of various political agents in Covid-related newspaper articles and tweets are sentiment-laden, ii) what are salient topics of criticism, and iii) how this evolved throughout January 2020 to April 2022. My analysis relies on a wide range of programming languages (including Python, HTML, CSS, and Javascript), tools (including Snorkel), and data science methods (including machine and deep learning, timeseries analysis, and topic modeling). With an accuracy of over 80%, I generate reliable findings on the status quo of political sentiment in the Swiss media. Most importantly, my work allows to deeper understand recent events, such as the pro-government reporting scandal faced by *Ringier* [71]. As a consequence, my analysis makes important sociopolitical and company-specific contributions. For example, it combats misinformation arising from undetected media bias by giving an analytically sound overview of the political leaning of newspapers. Moreover, the company I collaborated with, namely the *Neue Zürcher Zeitung (NZZ)*, has leveraged my work to publish an article on the reporting of Swiss newspapers throughout Covid-19.

Keywords: *Aspect-based sentiment analysis, weak labeling, Snorkel, BERT, topic modeling, non-negative matrix factorization , timeseries analysis, Box kernel, LOWESS*

Contents

1 Introduction

The Swiss population has shown a relatively government-critical reaction in the wake of Covid-19 and associated public health measures in comparison to neighboring countries, such as France, Italy, and Germany. According to a survey conducted in February 2022 by the *SRG*, the share of respondents placing great or very great trust in the Federal Council's pandemic-related measures declined from 61% in March 2022 to 45% in February 2022 [70]. As a consequence, the election surrounding the Covid-19 act in November 2021, which allowed the Federal Council to maintain necessary measures to manage the pandemic, generated a historically high voter turnout. A frequent point of criticism from opponents of the Covid-19 act and other measures associated with Covid-19 is the pro-government, unilateral, and uncritical coverage of the pandemic by the Swiss media. As such, only 24% of the respondents of the *SRG* survey believe that the media provides an objective account of the developments during Covid-19 [70].

Indeed, it has been recognized that appropriately covering the developments associated with Covid-19 poses a non-trivial challenge for media outlets, given the uncertainty surrounding the nature of the disease, the rapidly evolving conditions, and the overwhelming amount of information disseminated by various channels, termed as infodemic by the *World Health Organization (WHO)*. However, even in such circumstances, newspapers and other media outlets hold a central position within democratic processes by acting as public opinion makers. Through news and commentary sections, they provide readers with reliable information and a range of informed opinions pertinent to current issues. Thus, newspapers enable citizens to engage in public debate in an informed manner.

Next to newspapers, Twitter, too, has been recognized as a platform that is key to shaping public opinion [42] [26]. Twitter differentiates itself from other social media through its open platform design, where the user activity is public, as well as its micro-blogging concept, with a focus on succinct text-based posts. Thus, Twitter is more suitable for natural language processing analyses than other private, visuals-based platforms, such as Instagram or Facebook. While Twitter cannot be considered representative of the general public [52], it nonetheless forms a central online avenue for information gathering, opinion formation, and persuasion. Thus, Twitter forms a key arena for online public debate and studying this platform can reveal interesting insights into the dynamics of such discourses.

In light of the Swiss population's critical perception of the Covid-19 coverage in domestic newspapers and the integral role of newspapers with regards to public opinion, this project seeks to assess, quantify, and nuance the impartiality of the Covid-19 coverage of Swiss newspapers. Using natural language processing methods, the political sentiment in newspaper articles and tweets is investigated, allowing for a comparison between traditional and social media as key sources of public opinion. In particular, it is analyzed i) to what extent the mentions of various political agents in Covid-related newspaper articles and tweets are sentiment-laden, ii) what are salient topics of criticism, and iii) how this evolved throughout January 2020 to April 2022. The agents analyzed in this context include politicians, political bodies, institutions, officials, parties, and political camps.

This project was realized in collaboration with the *Neue Zürcher Zeitung (NZZ)*, one of Switzerland's leading newspapers.

In the following, I lay out the process behind and the results of this analysis. I give a summary of related previous work in Chapter 2. Here, I show how my analysis addresses a gap in the existing literature, as there have been no previous data-science-based analyses which aim to investigate the sentiment expressed towards political agents throughout the Covid-19 pandemic. In Chapter 3, I outline the data (Chapter 3.1) and explain the methodology underlying this project step by step (Chapter 3.2). In this section, I demonstrate that this analysis draws on a wide range of programming languages (including Python, HTML, CSS, and Javascript), tools (including Snorkel), and data science methods (including machine learning, deep learning, timeseries analysis, topic modeling, and data visualization). I discuss key results of my work in Chapter 4. Here, I provide details surrounding the performance of the supervised sentiment analysis (Chapter 4.1). Moreover, I present interesting insights regarding the political sentiment throughout the pandemic in Switzerland which can be gleaned from this analysis (Chapter 4.2). I also highlight the sociopolitical value of this project, which has led the *NZZ* to publish my work in a dedicated article in August 2022 (Chapter 4.3). Finally, in Chapter 5, I conclude my analysis, describe limitations at this stage, and consider potential future work. The Appendix in Chapter 6 contains an overview of the datasets used.

2 Literature review

In this Chapter, I give a summary of how my project relates to various previous analyses. I group existing literature into five clusters:

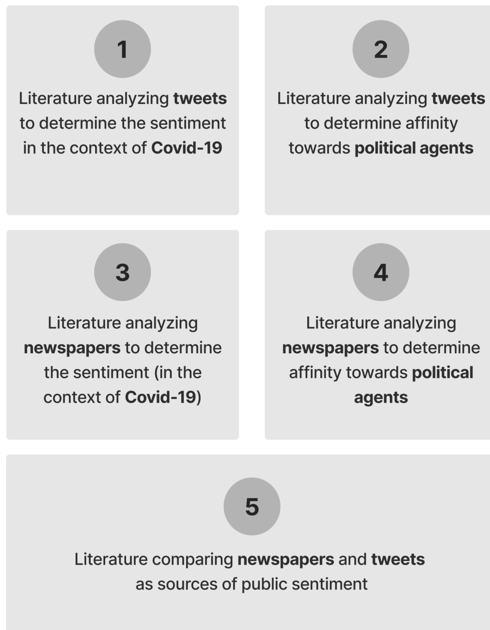

Figure 1: **Existing literature:** Previous work can be grouped into five clusters.

Various analyses have sought to capture the public opinion on Covid-19 through tweets (cluster 1). For example, Boon-Itt et al. [9], Chakraborty et al. [14], Imran et al. [37], Samuel et al. [65], Garcia et al. [27], Rustam et al. [64], and Shofiya et al. [67] have conducted analyses on Covid-19-related tweets to determine sentiment and recurring topics. The authors frequently relied on weak labeling via lexicon-based libraries, such as `AFINN`, `NRC`, `Sentimentr`, `SentiStrength`, `SentiWordNet`, `Syuzhet`, and `TextBlob`. The weakly labeled tweets were then classified using various supervised machine learning and deep learning models. Alternatively, the authors relied on wholly unsupervised transfer learning via pretrained, frequently BERT-based models.
In Switzerland, the *Digital Democracy Lab* [43] and Gilardi et al. [30] have analyzed the volume and development of Covid-19-related tweets from Swiss politicians, albeit without a sentiment aspect.

Similarly, a number of analyses have aimed to investigate affinity towards political agents in or through tweets (cluster 2). For example, Conover et al. [17], Boutet et al. [11], Cohen et al. [16], Barbera [6], Wong et al. [78], and Khatua et al. [40] have sought to classify the political orientation of Twitter users via

machine learning and deep learning models on the basis of features such as tweet content, retweet behavior, and social network structures. Bermingham et al. [7] and Ansaria et al. [2] have attempted to predict election results via machine learning and deep learning models using politician- or party-related tweet volumes and tweet sentiment.
In Switzerland, Müller et al. have made efforts to nuance the validity of Twitter as a data source for public opinion by establishing a dataset of Twitter accounts affiliated with a party [53].

Most sentiment analysis work has centered around obviously subjective text types, such as tweets and reviews. Given that newspaper articles convey opinion in a more implicit and nuanced manner, sentiment analysis targeting this source poses unique difficulties and requirements (see Chapter 3.2.3) [5].
Despite these intricacies, some analyses have been conducted which seek to analyze sentiment within the newspaper coverage (cluster 3). Outside of the Covid-19 domain, Remus et al. have explored sentiment within the German newspaper *Die Süddeutsche* as well as financial trading blogs using a proprietary lexicon and have related this to the developments of the DAX [61]. Oelke et al. have investigated polarity surrounding the mentions of soccer teams in various German newspapers via manual labeling and the `SentiWS` lexicon [55]. Lüdke et al. have established a GitHub repository to evaluate the sentiment development surrounding mentions of migration in German newspapers [49]. Hossain et al. have evaluated sentiment within the headlines of the Bangladesh newspaper *The Daily Star* using the `NRC` and `BING` lexica [35]. Lastly, Dehler-Holland et al. have conducted topic modeling and sentiment analysis surrounding the mentions of windpower in various German newspapers via the `SentiWS` lexicon [19].
For Swiss newspapers, Giehl and Mascarell et al. have explored sentiment, stance, and emotion. Giehl has developed a live web application visualizing sentiment for Swiss newspaper articles via manual labeling and the `TextBlob` library [29]. Mascarell et al. have established the CHeeSE dataset, consisting of debate questions and articles of Swiss newspapers which were annotated for stance and emotion classification [51].
Within the Covid-19 domain, Amazon Web Services has developed a Covid-19 news sentiment analyzer [1]. Costola et al. have investigated polarity surrounding the Covid-19 coverage of financial news platforms and have related this to stock market returns via transfer learning using a pretrained, BERT-based model [18]. Aslam et al. have evaluated the sentiment of Covid-19-related headlines for various English newspapers via the `Sentimentr` and `NRC` lexica [4]. Finally, MacKay et al. have explored the engagement rates and sentiment for Covid-19-related Facebook posts by newspapers and the federal public health department in Canada using the `SentiStrength` lexicon [50].

Furthermore, a number of analyses have been conducted which seek to investigate affinity towards political agents within the newspaper coverage (cluster 4). For example, Ansolabehere et. al [3] and Ho et al. [33] have provided and validated estimates of the political orientation of the editorial pages for U.S. newspapers. Gentzkow et al. have investigated political leaning as the similarity of a newspaper's language to that of a congressional Republican or Democrat [28]. Kaya et al. have performed a sentiment analysis on the mentions of various politicians, parties, and popular topics in Turkish newspaper columns. They employed various supervised classifiers, such as SVM and Naive Bayes [39], on a manually annotated dataset. Falck et al. have developed the Sentiment Political Compass, which provides a framework to analyze newspapers with regard to their political conviction. They conducted a sentiment analysis on the mentions of various politicians and parties in German newspapers using IBM Watson and, on this basis, mapped the newspapers onto the political compass [23].

One closely related analysis at the intersection of clusters 3 and 4, which investigates the sentiment of Swiss media outlets towards the Swiss government and public authorities during Covid-19, stems from the *Forschungszentrum Öffentlichkeit und Gesellschaft (FÖG)*. This research group periodically investigates the quality of coverage for Swiss media outlets. In 2020 and 2021, they published a special edition focusing on the Covid-19-related coverage. In this edition, various dimensions of media coverage were investigated, including the affinity towards the government and public authorities. The analysis relied on a manually labeled subsample of articles [21] [22].

Lastly, some analyses have been geared at comparing newspapers and Twitter in the context of Covid-19 or political affinity (cluster 5). For example, Gilardi et al. have performed topic modeling on Covid-19-related Swiss newspaper articles as well as tweets and compared the salience of different topics on social media versus traditional media [30]. Moreover, Li has investigated the sentiment surrounding mentions of parties in British newspapers and tweets using the `NRC` lexicon. These insights were contrasted with opinion polls in a correlation analysis [45].

To the best of my knowledge, there have been no previous data-science-based analyses which aim to investigate the sentiment expressed towards political agents throughout the Covid-19 pandemic in Switzerland or elsewhere. Giehl and Mascarell et al. have analyzed Swiss newspaper articles for sentiment, stance, and emotion, but lack a focus on political agents as well as Covid-19. Gilardi et al. have performed a comparative analysis of salient topics in Swiss newspaper articles and tweets related to Covid-19, but do not consider their sentiment. Lastly, the *FÖG* has conducted an analysis on the affinity of Swiss newspapers towards the government and public authorities throughout the Covid-19 coverage, but this analysis relied on manual labeling of a subsample of newspaper articles, did not assess sentiment for individual agents, and did not assess sentiment over the entire timeline of the pandemic [21] [22].
I aim to address this lacuna by investigating i) to what extent the mentions of various political agents in Swiss Covid-related newspaper articles and tweets are sentiment-laden, ii) what are salient topics of criticism, and iii) how this evolved throughout January 2020 to April 2022.

This project relies on various data science tools and models. The sentiment analysis was conducted using Snorkel for weakly labeling the dataset [38]. This approach was augmented by several models, including i) machine learning models, such as logistic regression [57], SVM [74], XGBoost [15], and Random Forest [12], ii) pretrained, BERT-based models [20], and iii) recurrent neural networks (RNNs) [66]. The topic modeling analysis was conducted using non-negative matrix factorization (NMF) [44]. A review of the background of these methods can be found throughout the methodology chapter.

3 Methodology

In this Chapter, I outline the data (Chapter 3.1) and explain the methodology underlying this project step by step (Chapter 3.2).

3.1 Datasets

In this project, two datasets were used, one of which contains tweets, the other of which contains relevant newspaper articles. The datasets were obtained by querying the Twitter and the *Swiss Media Database (SMD)* APIs for German-language, Covid-related documents between January 2020 and April 2022. For details on these queries and which variables the datasets include, see the Appendix Chapters 6.1 and 6.2.

3.2 Methodology

This project was realized using Python, HTML, CSS, and Javascript. It roughly followed five blocks, namely i) data gathering, cleaning, and preparation, ii) sentiment analysis, iii) timeseries analysis, iv) topic modeling, and v) data visualization:

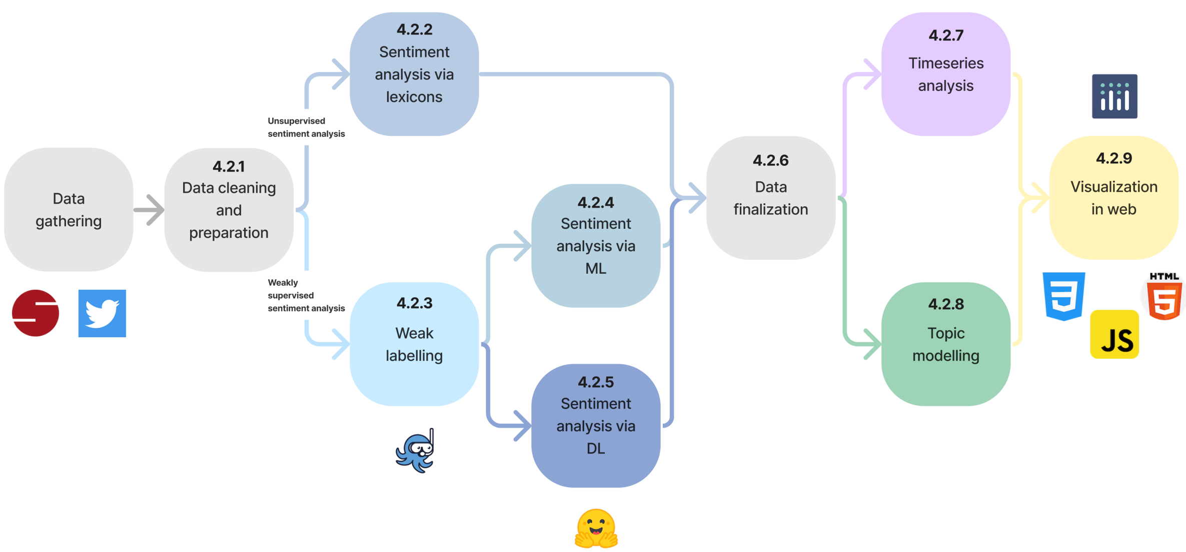

Figure 2: **Workflow and tech stack of the project:** The project proceeded in five main stages, namely i) data gathering, cleaning, and preparation (gray tones), ii) sentiment analysis (blue tones), iii) timeseries analysis (purple tones), iv) topic modeling (green tones), and v) data visualization (yellow tones). Blocks which are situated in parallel can be executed simultaneously, i.e. do not feature dependencies upon eachother. The numbering of the blocks refers to the chapters in which they are discussed.

For the sentiment analysis, this project relied on a weak labeling approach using Snorkel. The weak labeling approach was augmented by several machine-learning-based, transformer-based, and deep-learning-based methods as respective end models. A timeseries analysis of the sentiment was established using kernel-based smoothing methods and local linear regression. To identify polarizing topics, the sentiment analysis was followed by a topic modeling analysis using non-negative matrix factorization (NMF).
In the following, I give a detailed explanation of the methodology, both for the newspaper analysis and the Twitter analysis. Given that these rely on similar methods, I extensively describe my methodology for the newspaper analysis and only succinctly point out relevant differences for the Twitter analysis.

3.2.1 Data preparation

Prior to any sentiment analysis, both the newspaper dataset and the Twitter dataset were cleaned and structurally prepared.

Cleaning the newspaper dataset involved amending data types, dropping columns with a high share of missing values, dropping rows with impossible values (e.g. articles in English), formatting the article content (e.g. removing HTML tags), and conducting feature engineering (e.g. extracting the channel on which an article was published).

To capture the subtleties of opinion in newspaper articles, typically aspect-based sentiment analysis (ABSA) is adopted. ABSA serves to understand sentiment within documents as it relates to a given aspect. As an example, consider this sentence:

Insbesondere die SP hatte letzte Woche kritisiert, der Bundesrat habe bei den Sanktionen zu zögerlich gehandelt.
[Particularly the SP had criticized the Federal Council last week for its hesitation with regards to sanctions.].

This sentence contains two aspects, in this case the *SP [the Social Democratic Party of Switzerland]* as a party and the *Bundesrat [the Federal Council]* as a political body. While this sentence expresses a neutral sentiment towards the *SP*, it expresses a negative sentiment towards the *Bundesrat* as the recipient of criticism.
ABSA consists of three sub-tasks, namely i) extracting the aspects from the documents (aspect term extraction, ATE), ii) identifying opinion words relating to these aspects (opinion term extraction, OTE), and iii) classifying the sentiment for a given aspect (aspect term sentiment classification, ATC). While early approaches of ABSA treat these tasks as separate, recent approaches by Huang et al. [36], Peng et al. [58], and Shu et al. [68] have sought to combine these tasks into a single step.
ABSA is largely approached using neural networks, especially recurrent models [73], and transformer models, particularly BERT-based models [34]. Most models rely on some form of dependency parsing to determine the terms that are closely related to a given aspects. For BERT-based approaches, the task

is frequently framed as a question answering problem or sentence pair classification problem [34]. Other recent approaches include constructing a sentiment tree [10] or modeling the dependencies as a key-value memory network [73].

In this project, for improved comprehensibility, I adopted a three-stage ABSA approach where I treat ATE, OTE and ATC as separate tasks. To allow for such an analysis, the dataset required various structural amendments. First, as part of ATE, a number of political agents were located in the newspaper articles using regular expression search terms, one for each agent. One challenge that arose in this context was to identify the agents with a high precision and recall. As an example, consider the former Head of Communicable Diseases in the Federal Office of Public Health, Daniel Koch. Ideally, we want to identify all statements referring to Daniel Koch, even if they do so in different ways, such as *Daniel Koch*, *Koch*, or *Dr. Koch*. Furthermore, we want to exclude statements that do not refer to Daniel Koch, but other entities or concepts, such as *Robert Koch Institut [federal institution for disease monitoring and prevention in Germany]* or *kochen [cooking]*. To address this difficulty, the regular expressions were specified as precisely as possible, including defining casing, word boundaries, inflections, and lookaheads. Based on these regular expressions, relevant sentences in which an agent is mentioned were extracted. For a list of agents which were searched for and the corresponding regular expression search terms, see the Appendix Chapter 6.3.

Another challenge was posed by sentences, where multiple political agents are mentioned: For the sentiment analysis, it is important to identify only those words or subclauses relevant to a given agent, rather than base the sentiment analysis on the entire sentence. Hence, as part of OTE, sentences containing two distinct agents were transformed in two different fashions in order to capture the sentiment relating to a given agent more granularly. In a first approach, only those words which are grammatically closely related to the mention of an agent in the sentence were extracted. This was achieved by representing the syntactic child-ancestor connections between the tokens in a sentence as a network. Then, only those tokens were selected which were sufficiently close to the agent-token in the network. The measure of distance chosen was the normalized path distance, i.e. the number of tokens between a given token and the agent-token as a share of the total number of tokens in a sentence:

$$\text{Normalized path distance from token } i \text{ to the agent-token } a = \frac{\#\text{ of tokens between } i \text{ and } a}{\text{total } \#\text{ of tokens in the associated sentence } s}$$

It proved most beneficial to return tokens within a normalized path distance of 20% from the agent-token. Additionally, all adverbs and adjectives within a normalized path distance from the agent-token of 30% were extracted. As an example, let us consider the following sentence:

Als deswegen Epidemiologen den Bund kritisierten, tat sie Daniel Koch als Alarmisten ab.
[When epidemiologists criticized the Federal Council for this reason, Daniel Koch wrote them off as alarmists.]

The grammatical relationships in this sentence can be displayed as follows:

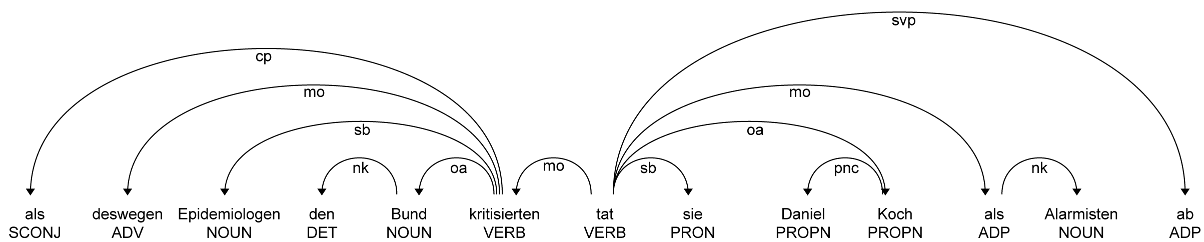

Figure 3: **Part-of-speech tags and syntactic relationships for example sentence:** The example sentence can be parsed into two subclauses.

For the agent *Daniel Koch*, the method identifies *kritisierten, tat, sie, Daniel Koch, als, Alarmisten, ab [criticized, wrote, them, Daniel Koch, as, alarmists, off]* as relevant tokens, as they lie within the maximum specified path distance from the agent-token.
In a second approach, only the subclause in which an agent appears was extracted. This was achieved by building a custom-made clause class which identifies subclauses based on the verb position. As an example, let us consider the following sentence:

Berger darf das, und politisch verantwortlich ist Berset.
[Berger is allowed to do so and Berset is politically responsible.]

For the agent *Berset*, the method identifies *politisch verantwortlich ist Berset [Berset is politically responsible]* as relevant subclause. Both approaches relied on `SpaCy's` dependency parser. Overall, the second approach yielded slightly more interpretable results and was therefore favored to parse sentences when multiple agents are mentioned. In contrast, if only one agent is mentioned, the sentence was considered in its entirety when performing the sentiment analysis.

The steps for cleaning and structurally preparing the Twitter dataset were largely the same as for the newspaper analysis. One interesting aspect were the emojis. One the one hand, emojis are a telling element when attempting to determine the sentiment of a tweet. On the other hand, in the context of Covid-19, many emojis are used that refer to illness, e.g. an emoji of a masked person. While these emojis normally hold a negative connotation, this may not necessarily be the case in the context of Covid-19. For this reason, two separate versions of the cleaned tweet content were generated. In one version, the emojis were entirely removed. In another version, the emojis were replaced by polarized keywords based on the `emosent-py` ranking [56], namely *fantastisch, gut,* *schlecht, schrecklich [fantastic, good, bad, terrible]* to account for their emotional content during the sentiment analysis.

The final newspaper dataset after data cleaning and parsing contained 268'001 mentions of political agents across 41'547 newspaper articles by 48 sources. The final Twitter dataset contained 257'497 tweets. For details on the variables recorded in these datasets, see the Appendix Chapters 6.4 and 6.5.

3.2.2 Proof-of-concept via lexicon-based approaches for sentiment analysis

To determine the sentiment surrounding the mentions of political agents in newspaper articles and tweets, I first deployed lexicon-based classifiers, including `TextBlob` [47] and `SentiWS` [62], as a proof-of-concept. These methods rely on dictionaries, where words are annotated with regards to the polarity of their connotation. They have the advantage that they can be deployed out-of-the-box. However, machine-learning-based and deep-learning-based classifiers generally outperform lexicon-based methods. Thus, they typically only serve as baseline models [24]. Moreover, rule-based classifiers are unsupervised methods. Hence, analyses relying on such methods may lack transparency regarding their reliability.

Prior to deploying the lexicon-based classifiers, the dataset was preprocessed. This included lowercasing of the article content, tokenization, stopword removal using `NLTK`, punctuation removal, and lemmatization using `SpaCy`.
`TextBlob` forms one of the most popular rule-based sentiment analysis libraries. It was released by Steven Loria in 2013. In 2014, a German language extension, `textblob-de`, was developed by Markus Killer [47]. I chose `TextBlob` as the first proof-of-concept classifier as it allows to extract both sentiment and subjectivity scores for a given document, whereas many other classifiers only allow to extract sentiment scores.
`SentiWS` was released by the University of Leipzig in 2012 [62]. I chose `SentiWS` as the second proof-of-concept classifier as it forms one of the most performant rule-based sentiment analysis libraries for the German language [24]. In contrast to `TextBlob`, `SentiWS` only allows to extract sentiment scores. The sentiment scores are given on a token-level. Following Fehle et al., the overall sentiment for a sentence was calculated as follows [24]:

$$S_s = t_p - t_n$$

$$S_c = \begin{cases} 1, & \text{if } S_s > 0 \\ -1, & \text{if } S_s < 0 \\ 0, & \text{otherwise} \end{cases}$$

Here S_s is the unnormalized sentiment score assigned to a sentence S, S_c is the sentiment class assigned to a sentence S, t_p is the number of positive tokens contained in S, and t_n is the number of negative tokens contained in S.

3.2.3 Generation of weak sentiment labels via Snorkel

To ensure transparency regarding the reliability of the sentiment analysis and improve classification performance, I followed the unsupervised proof-of-concept by a supervised approach. Given the size of the dataset, manually annotating the sentiment labels proved infeasible. Thus, I opted for a weak labeling approach.

Weak supervision approaches were developed in response to the bottleneck presented by manually labeling data. This task is time-consuming, expensive, and poses difficulties when the underlying data is updated. Weak supervision allows to rapidly create large training sets by relying on several weak supervision sources to generate labels, instead of manual labeling [79]. There are various types of weak supervision, including incomplete supervision, inexact supervision, and inaccurate supervision [80]. My approach falls within the incomplete supervision domain, where a small subset of the data is annotated, while the remaining data remains unannotated. The labels for the unannotated data are then generated by aggregating multiple sources of weak supervision.

For the newspaper dataset, 5% of the total agent mentions were manually annotated for sentiment, determining whether an agent mention is negative, positive, or neutral. For the larger Twitter dataset, only 2% of the total agent mentions were manually annotated. Given the subtleties of opinion in newspaper articles, this task merits further explanation. When conducting sentiment analysis in the context of newspaper articles, it is important distinguish along two axes, namely descriptive versus evaluative content and positive versus negative content. Newspapers frequently make descriptive statements, which may be perceived as positive (e.g. *A fleet of volunteer bus drivers is helping people escape eastern Ukraine*), neutral (e.g.*The war in Ukraine, as seen on Russian TV*), or negative states of the world (e.g. *Shelling hits a kindergarten in Ukraine*). More rarely, newspapers also make evaluative statements, which may express a positive (e.g. *An unlikely hero*[1]) or negative sentiment (e.g.*Inside Trump's failure: The rush to abandon leadership role on the virus*). The distinction between descriptive and evaluative statements is a matter of philosophical debate and may not always be straightforward in practice. As an example, consider the statement *Republicans wrongly blame Biden for rising gas prices.* It is not immediately clear whether this statement expresses mere facts or a negative evaluation of Republicans. In response, Balahur et al. suggest to treat evaluative statements as statements which are not obviously verifiable [5]. Furthermore, the distinction between positive and negative may not always be unambiguous in practice, depending on the worldview of the reader and author. As an example, consider the statement *Conservatives lose 2 races in U.K.* For a reader in favor of the conservative party, this statement may be perceived as negative, whereas it may be positive for a reader who vehemently opposes the conservative party. Therefore, Balahur et al. recommend to adopt a strictly text-oriented perspective as far as possible, refraining from judging polarity from the standpoint

[1]Referring to Volodymyr Zelensky.

of a particular reader or author and assuming a limited world knowledge [5]. Furthermore, I adopted the following annotation guidelines as far as possible:

1. Only evaluative statements, rather than descriptive statements, are annotated
2. Statements are annotated without reference to deep knowledge of the target agent and the newspaper's political orientation
3. Statements are annotated with regards to the sentiment expressed towards the target agent, rather than sentiments expressed by the target agent or towards other, unrelated circumstances
4. In case of conflicting sentiments within one statement, the statement is annotated as neutral

The gold labels obtained through manual annotation of 5% of the dataset were then used to generate silver labels for the remaining 95% of the dataset via weak supervision with Snorkel [38].
Early examples of weak supervision include crowdsourced labels or rule-based heuristics. However, it was soon recognized that aggregating several weak supervision sources serves to increase the accuracy and coverage of the labels [60]. Challenges arising in this process are that, firstly, the sources may overlap and generate conflicting labels, and that, secondly, resolving conflicting labels requires estimation of the accuracies and correlations among the sources.
In response, Ratner et al. developed Snorkel as a tool to generate a single, probabilistic label for a given datapoint by combining various, user-defined weak supervision sources [59]. Snorkel's workflow proceeds in three stages:

1. **Define labeling functions:** The user defines several labeling functions, which either output a label for a given datapoint based on a certain source of weak supervision or abstain. These sources can range from crowdsourced labels over rule-based heuristics to machine learning approaches, including unsupervised models, supervised models trained on a small, manually labeled training set, and transfer learning via pretrained, supervised models.
2. **Model accuracies and correlations among labeling functions:** Based on the agreements and disagreements of the labeling functions, Snorkel learns a generative label model, allowing it to estimate the accuracies and correlations among the labeling functions. It, then, outputs a single, probabilistic label for a given datapoint. This label represents a weighted combination of the labels output by the labeling functions.
3. **Model final label:** Based on the labels generated through the generative model, a discriminative end model is learned, which may allow to improve generalization beyond the labeling functions. Typically, the stages of learning the label model and the end model are decoupled in a two-stage process [79].

Expressed mathematically, the aim is to learn a classification model h_θ that, given a datapoint $\mathbf{x} \in \mathcal{X}$, predicts its label $y \in \mathcal{Y}$.
The labeling functions, $\lambda : \mathcal{X} \rightarrow \mathcal{Y} \cup \{\varnothing\}$, either output a label for a given datapoint or abstain, the latter being denoted by $\varnothing$. Given m datapoints and n labeling functions, Snorkel produces a matrix of labeling function outputs, $\mathbf{\Lambda} \in (\mathcal{Y} \cup \{\varnothing\})^{m \times n}$.
Snorkel, then, aggregates this matrix, which may contain overlapping and conflicting labels, into a single vector, $\mathbf{y} = (y_1, ..., y_m)$ through a generative model, g_w. Generative models form part of probabilistic classification paradigms, where the aim is to learn the posterior, $p_\mathcal{Y}(y|\mathbf{x})$. In the generative approach, $p_\mathcal{Y}(y|\mathbf{x})$ is learned indirectly through the likelihood $p_\mathcal{X}(\mathbf{x}|y)$ and $p_\mathcal{X}(\mathbf{x})$. In aggregating the labeling function matrix, Snorkel generates a concatenated vector for each datapoint, $\boldsymbol{\phi_i}$, which includes the label propensity and accuracy for each labeling function as well as the correlations for each labeling function pair. This vector forms the input for the generative model, which is optimized upon the parameters, $\mathbf{w}$:

$$g_w = \text{constant} \times \exp(\sum_{i=1}^{m} \mathbf{w}^T \boldsymbol{\phi_i})$$

The parameters are optimized by minimizing the negative log marginal likelihood given the observed label matrix $\mathbf{\Lambda}$:

$$\tilde{\mathbf{w}} = \arg\min_{\mathbf{w}} \ -\log \sum_{Y} g_w$$

Given the parameters, the generative model outputs labels $\tilde{Y} = g_{\tilde{w}}(Y|\mathbf{\Lambda})$. The labels resulting from the generative model, $\tilde{Y}$, are then used to train a discriminative model, h_θ. Discriminative models, too, form part of probabilistic classification paradigms. In contrast to the generative approach, $p_\mathcal{Y}(y|\mathbf{x})$ is learned directly. Discriminative models rely on fewer assumptions and are therefore more robust than generative models. The parameters of the discriminative model are learned by minimizing a noise-aware variant of the loss, $l(h_\theta(\boldsymbol{x_i}), y)$ [59]:

$$\tilde{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \ \sum_{i=1}^{m} \mathbb{E}\left[l(h_\theta(\boldsymbol{x_i}), y)\right]$$

In this project, the labeling functions included lexicon-based, keyword-based, machine-learning-based, and deep-learning-based functions. As lexicon-based methods, I employed `TextBlob` and `SentiWS`.
Keyword-based methods output a sentiment for a statement based whether they contain certain word forms. For example, I defined a function to return a negative sentiment, when the statement involves a common hashtag expressing a negative stance, such as *#nichtmeinbr [#notmyfederalcouncil]*, *#swisscovidfail*, and *#swisscovidcrime*.
Machine-learning-based methods refer to supervised classification models, such as logistic regression, support vector machines (SVM), XGBoost, and Random

Forests. I trained these models using bag-of-words (BOW), term-frequency inverse-document-frequency (TF-IDF) vectorization, and various pretrained embeddings, such as `fastText, Spacy, GloVe,` and `Word2Vec`.
Lastly, deep-learning-based methods refer to pretrained, frequently transformer models, such as BERT. I used models trained by Guhr [32] and Lüdke et al. [48].

The labels output by the labeling functions are then aggregated to a single silver label per datapoint by Snorkel. One challenge encountered in this process was that the precision for positive statements was very low (see Chapters 4.1.1 and 4.2.1). For this reason, I manually re-coded the default Snorkel label model to return a positive label only if at least three labeling functions output a positive label and, otherwise, return a neutral label.
Despite my adjusted model, the labeling functions performed comparably poorly for positive statements, driven by the circumstance that these formed merely 4-5% of the manually labeled data for both the newspaper dataset and the Twitter dataset (see Chapters 4.1.1 and 4.2.1). For this reason, the further analyses and visualizations were reduced to a binary classification problem, distinguishing between negative and non-negative statements only.

3.2.4 Machine learning models for sentiment analysis

Having established a weakly labeled dataset, I continued with a supervised classification approach for sentiment analysis, predicting whether an agent mention is negative or non-negative. As alluded to previously, the aim was to obtain a standalone end model that generalizes better than the labeling functions, as these were trained only on a small-scale, manually annotated dataset, whereas the machine learning models could now be trained on a large-scale, weakly annotated dataset.

Prior to model training, two different preprocessing methods were applied to the data, namely vectorization, such as BOW or TF-IDF, and pretrained embeddings, such as `fastText, Spacy, GloVe,` or `Word2Vec`. The aim was to select the best vectorization technique and the best embedding for further fine-tuning of the machine learning models. BOW and TF-IDF are preprocessing techniques that allow to represent documents, in this case sentences or tweets, as vectors based on the vocabulary of the corpus. Under BOW, a document is represented as a sparse vector $\mathbf{v}$, where each dimension v_i represents a term of the vocabulary and takes a value based on the number of times this term appears in the document. Under TF-IDF, a document is also represented as a sparse vector, but the value for v_i is calculated based on the term frequency, TF, and the inverse document frequency, IDF:

$$TF(t,d) = \frac{\text{number of times t appears in d}}{\text{total number of terms in d}}$$

$$IDF(t) = \log \times \frac{\text{total number of documents in c}}{1 + \text{total number of documents containing t}}$$

$$TF\text{-}IDF(t, d) = TF \times IDF$$

Here, t is a given term, d is a given document, and c is the corpus.
Another option besides BOW and TF-IDF are pretrained embeddings, where the tokens contained in a document are represented as pretrained, fixed-size vectors. The document can, then, be represented as the mean of these token vectors. Pre-trained embeddings offer the advantage that they represent the syntactic and semantic similarities between terms in a high-dimensional vector space and, hence, allow for a richer understanding of documents.
To determine a baseline, each machine learning model was initially tested under each vectorization method and embedding without any finetuning. When comparing BOW versus TF-IDF vectorization, almost all models showed better performance with a simple BOW vectorization. When comparing the embeddings, all models showed strong performance with Spacy vectors. Given the lack of a library for GloVe and Word2Vec embeddings in the German language, these embedding methods were not considered further. Thus, each model was further finetuned once on a BOW-vectorized dataset and once on the Spacy embeddings of the dataset.

As machine learning models, I tested a logistic regression, SVM, XGBoost, and Random Forest. Logistic regression is a classifier that fits a linear hyperplane between the classes. It can be motivated by appeal to an additive logistic noise latent variable model, where we assume that the data is generated by a latent variable $\mathcal{Y}$ and that the noise is distributed according to a logistic distribution. Given the simplicity associated with the linear discriminant, it generally serves as a baseline classifier [57].
SVM is a classifier that fits a hyperplane with the widest possible margin between the classes, as this is linked to better generalization properties. In soft-margin SVM, a tolerance for margin errors is established, meaning not all datapoints must strictly lie within their class area. Given the mathematical properties of SVM, it qualifies for the kernel trick. Kernels provide a mechanism to map data into a higher-dimensionsal feature space and employ linear models within this feature space. Thus, SVM can fit even highly non-linear data well [74].
XGBoost is a tree-based ensemble method that uses a gradient boosting framework. The classifier is trained sequentially and each classifier attempts to improve its predecessor. This is achieved by inputting the residuals of the previous classifier as label for the training data. In this way, the algorithm is directed to focus on datapoints, where the previous classifier failed to correctly predict the label [15].
A Random Forest is also a tree-based ensemble method. It combines multiple decision trees, where each training set is randomly generated via bootstrapping. At each split within the decision tree, a randomly bootstrapped fraction (approximately $\sqrt{k}$) of the features is considered [12].

These four models were finetuned with regards to their various hyperparameters.

I selected the F1 macro score as metric of evaluation.

$$F1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

$$= \frac{\text{True positives}}{\text{True positives} + \frac{1}{2} \times (\text{False positives} + \text{False negatives})}$$

Here, precision refers to the ability of a model to identify only the relevant data points, at the cost of forgoing some true positives. Recall refers to the ability of a model to identify all the relevant data points, at the cost of generating some false positives.
Relying on the F1 score allows to optimize both upon precision and recall, as the F1 score combines these two metrics. Moreover, relying on a macro average mitigates negative implications arising from the imbalance between the classes, as the F1 scores are averaged without consideration of the proportion for each label in the dataset.
To finetune the models, I employed a greedy, iterative grid search approach. In this approach, I tune key hyperparameters step by step. Starting with one hyperparameter, I first select common hyperparameter values based on industry standards. Then, I adjust and optimize these values in a grid search. Once I find the optimum value for a given hyperparameter, it is passed as a fixed hyperparameter to the next gridsearch, where I tune the next hyperparameter. While this approach is not guaranteed to find the best model, it is computationally efficient and likely to find a reasonably good model.

As a final model, an ensemble voting model was tested, combining the logistic regression, SVM, XGBoost, and Random Forest. Ensemble methods serve to combine several standalone models by taking a majority vote across their predictions. This typically works best when the standalone models are structurally different from each other.

3.2.5 Deep-learning-based models for sentiment analysis

Aside from machine-learning-based models, I also tested a deep-learning-based classification approach for sentiment analysis. I trialled three approaches, including i) transfer learning with four pretrained, BERT-based models, ii) manual finetuning of the best-performing pretrained model, and iii) building a `Keras`-based neural network with embeddings from scratch.

For the transfer learning approach (i), I tested four BERT-based models. Bidirectional encoder representations from transformers (BERT) are a transformed-based models. They rely on a encoder-decoder structure with self-attention. Encoder-decoder models are composed of a sequence-to-vector network, the encoder, followed by a vector-to-sequence network, the decoder. The encoder is responsible for reading and encoding the input sequence. It does this in a bidirectional fashion, instead of a left-to-right and/or right-to-left fashion. When seeking to encode a given word, the encoder can consider other positions within

the same sequence. This is called self-attention. The decoder is then responsible for generating the prediction from the encoded sequence [20].
Two of the models I trialled are classic pretrained, BERT-based sentiment classification models, classifying documents as negative, positive, or neutral. The model by Guhr was trained on 1.8 million German-language samples across various domains [32]. The model by Lüdke et al. is based on the model by Guhr, but was additionally trained on German newspaper articles [48].
The third model I tested is a pretrained, BERT-based review classification model established by NLP Town, classifying documents between one to five stars [75]. I manually re-coded this model for my purposes: One-star reviews were considered negative statements and five-star reviews were considered positive statements. All remaining reviews were classified as neutral.
The final model I trialled is `PYABSA`, a pretrained, DeBERTa-based aspect-based sentiment classification model, classifying statements regarding specified aspects as negative, positive, or neutral. It requires marking the aspects, in this case the political agents, with `[ASP]<agent>[ASP]`.
Both for the newspaper analysis and the Twitter analysis, the model that achieved the highest F1 macro score among these four pretrained models was selected for further finetuning.

Finetuning (ii) was achieved by feeding the outputs of the last hidden layer of the pretrained model into single-hidden-layer feed-forward BERT classifier. First, the inputs are preprocessed, converted to tensors, and packed into a dataloader for improved computational efficiency. During model training, a forward pass is performed to compute the logits and the training loss. Then, a backward pass is performed to compute the gradients. At this step, additional features, such as gradient clipping, are implemented to avoid common challenges, such as exploding gradients. Finally, the model's parameters and the learning rate are updated. During model evaluation, the process is similar. However, here only a forward pass is performed to compute the logits. Then, a softmax function is applied to determine the class probabilities. Furthermore, the validation loss and accuracy are computed.

Lastly, I built a `Keras`-based neural network with embeddings from scratch (iii). Neural networks typically consist of an input layer, several hidden layers, and an output layer. Weights and activation functions are applied to the input features in the input layer and the neurons in deeper layers. A special form of neural networks, which are used for the processing of sequences like texts, are recurrent neutral networks (RNN). Their hidden layer is formed by a memory cell, which allows for the persistence of information over time. By recursively stepping through the input sequence and relying on said memory cell, the RNN can utilize information from prior inputs to determine how to process the current input [66].
In my RNN, the inputs are first converted into integer sequences and padded. Then, a sequential model is instantiated, with a trainable Embedding layer, long short-term memory layer, GlobalMaxPooling1D layer, and several dense layers. This model is compiled using the Adam optimizer and its default hyperparam-

eters. This optimizer combines the benefits of momentum optimization, which increases the step size based on previous gradients, and RMSProp optimization, which scales the gradient vectors along its steepest dimensions to promote a straight path towards the optimum on a curved error surface. Finally, the model is trained with an early stopping monitor and a custom-defined decaying learning rate.

3.2.6 Dataset finalization

For the newspaper analysis, the weak sentiment labels were selected for further analysis and visualization, based on their superior classification performance versus the standalone machine learning and deep learning models. For the Twitter analysis, the labels output by the `Keras`-based neural network were selected. To prepare the data for the next steps, various additional variables were mapped to the newspapers and agents, such as publisher, party affiliation, etc.

3.2.7 Timeseries analysis for sentiment analysis

Given the sentiment-annotated, finalized dataset, I aggregated the average sentiment for selected political agents on a timeseries basis. The aim was to find a smoothing method which is suitable for irregular timeseries and visualization. To this end, I tested eight smoothing methods. First, I tested a simple moving average with a window size of 50 days. It is calculated as follows:

$$y_t = \frac{1}{s} \sum_{i=\frac{1-s}{2}}^{\frac{s-1}{2}} y_{t+i}$$

Here, t represents a point in time and s represents the window size.
Then, I tested two exponential smoothing methods, including a simple exponential moving average and a Holt-Winters exponential moving average. The simple exponential smoothing method takes a weighted average of the neighboring points within the window, where the weights exponentially decline towards the boundaries of the window. The Holt-Winters method additionally removes trends and seasonality in the data.
Based on the findings of Falck et al. who used a box kernel for timeseries smoothing [23], I also tested several kernel methods, including a Box1DKernel, a Gaussian1DKernel and a Trapezoid1DKernel. In each case, the kernel defines the precise shape of the function used to take the average of the neighboring datapoints within the window respectively the kernel width.
Lastly, I tested two smoothing methods, which are able to smooth out even high frequencies and generate jitter-free series, which is particularly suitable for visualization purposes. These methods are triangle smoothing and locally weighted scatterplot smoothing (LOWESS) smoothing. The triangle smoother relies on a weighted aggregation of the neighboring points within the window, where the weights linearly decline towards the boundaries of the window. In contrast, the

LOWESS smoother relies on a weighted linear regression of a fraction of the datapoints closest to a given value.
One challenge that emerged in this context was to find a stable, but nonetheless jitter-free smoothing method. To address this, I developed a combination of the Box1DKernel and the LOWESS smoother, where the timeseries is first smoothed via the kernel. Thereafter, any high frequencies are polished via LOWESS. This method provided the most visually interpretable results and was therefore selected for further analysis and visualization.

3.2.8 Topic modeling

To identify key points of criticism for the political agents, I tested three unsupervised topic modeling methods, namely i) Latent Dirichlet allocation (LDA), ii) k-means clustering, and iii) non-negative matrix factorization (NMF). The aim was to find a topic model that output comprehensible topics which aligned with well-known historic events.
While the sentiment analysis was conducted on a sentence-level, identifying topics on a passage-level proved more interpretable, given the additional context provided. Prior to topic modeling, the passages were preprocessed. This included tokenization, stopword removal using `NLTK`, punctuation removal, agent name removal, and lemmatization using `SpaCy`.

LDA (i) is a topic modeling technique, which assumes that each document can be described by a distribution of topics and that each topic can be described by a distribution of terms. However, the topics are are latent and only the documents and terms are observable. LDA proceeds as follows: First, it assumes there are k topics across the documents. It is assumed that these topics are distributed across the document d according to the Dirichlet distribution α. Moreover, it is assumed that the terms are distributed across a topic t according to the Dirichlet distribution β. Then, these two sets of probabilities are computed using Bayesian methods and an expectation maximization algorithm [8].
In this project, LDA was fit on a TF-IDF vectorized corpus for a given agent, with alpha set to asymmetric, beta set to auto, and the decay set to 0.6. The alpha hyperparameter controls the prior for the document-topic distribution, whereas the beta hyperparameter controls the prior for the topic-word distribution. The decay controls the learning rate. For each agent, three topics were identified. These parameters were determined based on tuning the perplexity and coherence for LDA. Perplexity indicates how well the model represents the statistics of a held-out dataset. A lower perplexity indicates better generalization ability. Coherence measures the degree of semantic similarity between high scoring words in the topic. A higher coherence measure indicates greater interpretability.

K-means clustering (ii) is a partitional clustering algorithm. It relies on a predefined number of clusters and clusters datapoints based on their interpoint proximity [54].

In this project, I fit the clusters on pretrained `Spacy` embeddings of the passages for a given agent. Then, based on TF-IDF, the most common words were identified for each cluster. For each agent, three clusters and, thus, topics were identified.

NMF (iii) is a dimensionality reduction technique, but can also be employed for topic modeling. As an input, it takes a term-document matrix $\mathbf{A}^{n \times m}$, either based on BOW or TF-IDF. It outputs two non-negative matrices, one of which is a term-topic matrix $\mathbf{W}^{n \times k}$, the other of which is a topic-document matrix $\mathbf{H}^{k \times m}$. Here, n is the number of terms in the vocabulary, m is the number of documents, and k is the number of topics. These matrices $\mathbf{W}$ and $\mathbf{H}$ are found by optimizing upon the following equation: $\mathbf{A} = \mathbf{W} \times \mathbf{H}$ [44].
In this case, NMF generated the most comprehensible results. It was fit on a TF-IDF vectorized corpus for a given agent. The parameters were tuned for each agent individually. Alpha was set within the range of 0.075 to 0.2. The $\ell 1$ ratio was set within the range of 0.1 to 0.7. Both alpha and the $\ell 1$ ratio are regularization hyperparameters. The number of topics identified per agent ranged from 3 to 7.

3.2.9 Visualization and webpage

The visualizations of the sentiment and topic modeling analyses were realized in `Plotly` to ensure interactivity.
One challenge was visualizing the results of the topic modeling analysis. Here, I tested two approaches. In a first approach, I reduced the dimensionality of the dataset generated by NMF, where all words in the corpus are listed alongside the strength of their association with the top three topics for a given agent. To this end, I used a t-distributed stochastic neighbor embedding (tSNE). This is a non-linear dimensionality reduction technique. It operates in three stages: In a first step, tSNE constructs a probability distribution over pairs of high-dimensional datapoints in such a way that similar datapoints are assigned a higher probability. In a second step, it constructs another probability distribution over the points in the low-dimensional map. Finally, tSNE minimizes the Kullback–Leibler divergence between the two probability distributions with respect to the locations of the points in the low-dimensional map [77].
In this way, for each word in the corpus, the strength of its association with the three topics was mapped to two dimensions. Then, the words were visualized a scatterplot and colored according to their topic. This visualization allows readers to explore top criticism keywords for an agent, understand the "span" of a topic of criticism, i.e. how many words are associated with it, and understand the "coherence" of a topic of criticism, i.e. how closely related the keywords within a topic of criticism are amongst each other. As an example, see Figure 4.

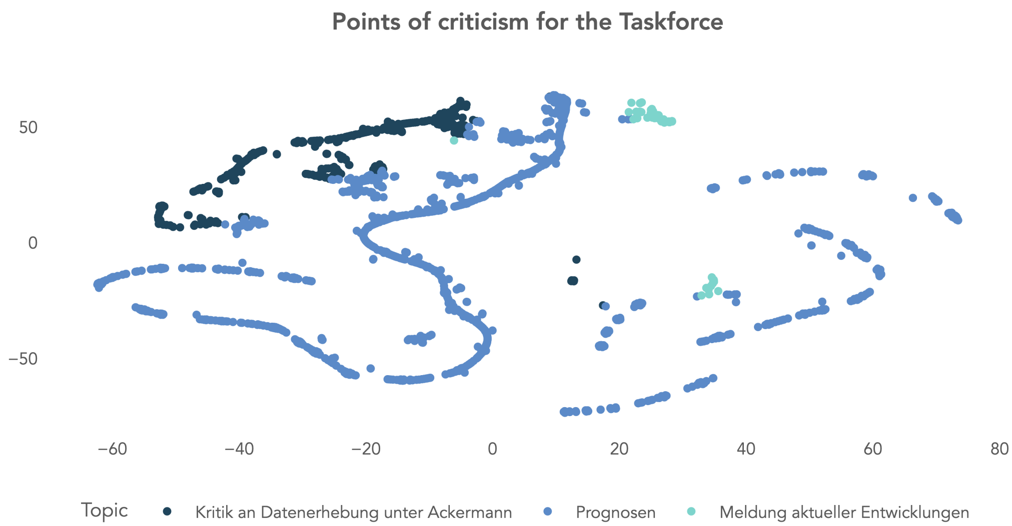

Figure 4: **Scatterplot visualization of topic model for the *Taskforce*:** For the *Taskforce*, three topics were identified. Each scatterpoints represents a key word from the corpus associated with a given topic.

In a second approach, I visualized the topics and the words pertaining to them as a network. Each word was connected to its topic via an edge weighted by the strength of its association with the topic, if the latter exceeded 0.05. Furthermore, words were connected via edges amongst each other, if both words shared an association with a topic with a strength of at least 0.15. As an example, see Figure 5.

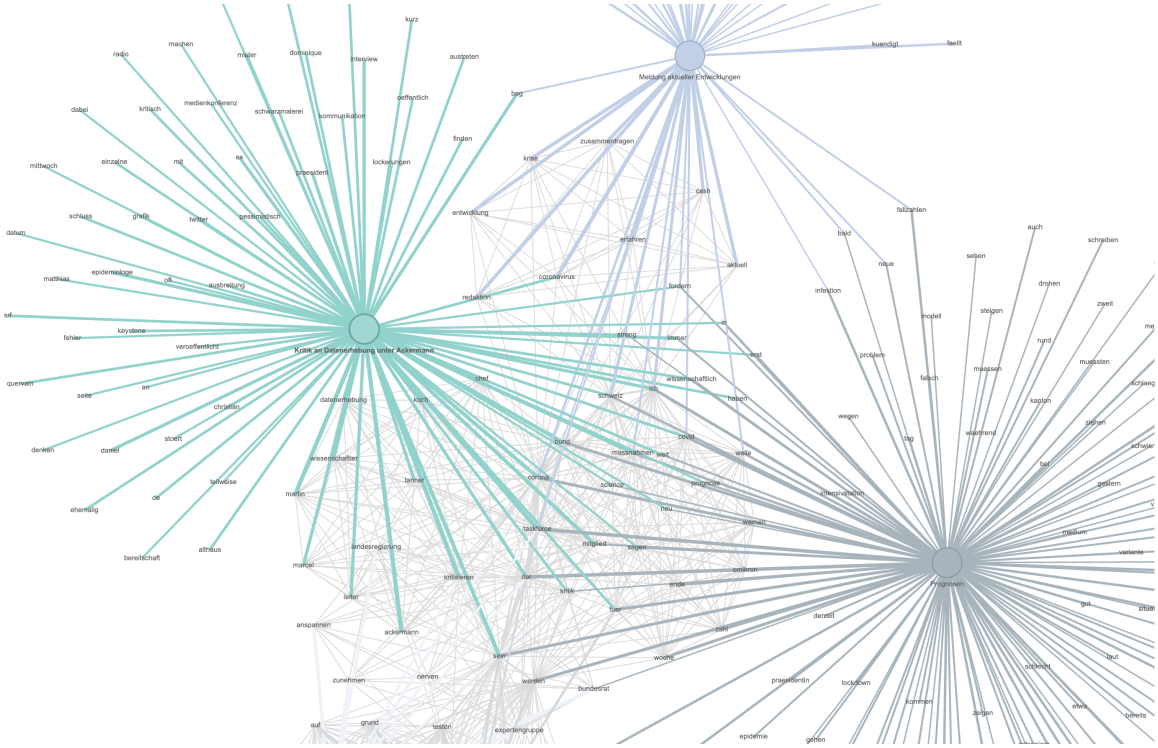

Figure 5: **Network visualization of topic model for the *Taskforce*:** For the *Taskforce*, three topics were identified, represented by the larger nodes. Each smaller node represents a key word from the corpus associated with a given topic. Words, which are highly interconnected, also share an edge.

To offer readers a platform to view key insights, I opted for a webpage. On the webpage, the reader is provided with interactive visualizations alongside a blog-form interpretation of key results. I developed the webpage using HTML, CSS, and Javascript for the frond-end. While HTML structures the different subpages, CSS determines a consistent style across the subpages, and Javascript renders the visualizations interactive.

4 Results

In this Chapter, I discuss key results of my work. I first provide details surrounding the performance of the supervised sentiment analysis (Chapter 4.1). This includes the performance of the weak labeling, the machine learning models, and the deep learning models. I then present interesting insights regarding the political sentiment throughout the pandemic in Switzerland (Chapter 4.2). Here, I both describe which insights can be gleaned from my analysis and how historic events render my analysis more plausible. I then highlight the sociopolitical, company-specific, and technical contributions which my project makes (Chapter 4.3). I also refer to a webpage with an interactive presentation of the insights from my analysis (Chapter 4.4).

4.1 Supervised sentiment analysis performance

4.1.1 Newspaper analysis

Weak labeling: To weakly label the newspaper dataset, I established twelve Snorkel labeling functions with an accuracy of over 50%. Here, accuracy refers to the share of predictions correctly labelled by the function, when applied to the manually labelled test set.

LF	**Coverage**	**Conflicts**	**Accuracy**
textblob	1.0	0.73	0.66
sentiws	1.0	0.73	0.70
polart	1.0	0.73	0.59
gervader	1.0	0.73	0.63
claim	0.12	0.09	0.89
log_reg_cv_lf	1.0	0.73	0.94
svm_cv_lf	1.0	0.73	0.95
xgbr_cv_lf	1.0	0.73	0.94
rf_cv_lf	1.0	0.73	0.92
xgbr_spacy_lf	1.0	0.73	0.69
rf_spacy_lf	1.0	0.73	0.81
bert_mdraw_lf	1.0	0.73	0.74

Table 1: **Performance of the Snorkel labeling functions:** This table shows all labeling functions, including the share of documents they apply to (respectively do not abstain from), the share of other labeling functions they are in conflict with, and their accuracy. A SVM model trained on count-vectorized statements shows the strongest performance.

The labels output by the individual labeling functions were, then, aggregated into a single silver label per datapoint via the Snorkel label model. My customized label model, which is particularly careful when labeling statements as positive, achieved the following performance when applied to the manually labelled test set:

Label	**Precision**	**Recall**	**F1**	**Support**
Neutral	0.95	0.95	0.95	1976
Positive	0.49	0.31	0.38	124
Negative	0.70	0.81	0.75	340
Accuracy			0.90	2440
Macro average	0.71	0.69	0.70	2440

Table 2: **Performance of the Snorkel labeling functions:** While the performance for neutral and negative statements is sufficiently strong, the label model frequently fail to correctly identify all positive statements, indicated by a low recall.

The silver labels output by the label model featured a weighted average accuracy of 90%. Both precision and recall were sufficiently high for neutral and negative statements. For positive statements, however, the label model frequently failed to correctly identify all relevant statements, indicated by a low recall. As described in Chapter 3.2.3, the problem was subsequently framed as a binary classification problem, distinguishing between negative and non-negative statements only.

Machine learning models: Among the standalone machine learning models, the ensemble voting model generated the best result in terms of the F1 macro:

Label	**Precision**	**Recall**	**F1**	**Support**
Non-negative	0.90	0.97	0.94	2123
Negative	0.49	0.19	0.27	282
Accuracy			0.88	2405

Macro average	0.69	0.58	0.60	2405

Table 3: **Performance of the ensemble model on gold labels:** The ensemble model achieves an accuracy of 88%.

When tested on the gold and silver labels combined, the ensemble model achieved a weighted average accuracy of 93%. When tested only on the gold labels, it featured a weighted average accuracy of 88%. For negative statements, the recall was significantly lower than the precision. This indicates that, while the statements labelled as negative by the models were likely to be true negatives, the model missed out on a large number of negative statements, which were falsely labelled as neutral.

Deep learning models: Among the four pretrained models, the model by Lüdcke et al. performed best with a weighted average accuracy of 80%, with a F1 score of 89% for neutral statements and 32% for negative statements. In contrast, `PYABSA` showed relatively weak performance with an F1 score of only 8% for negative statements. This may be driven by the circumstance that ABSA is generally a highly domain-dependent task [68] and that the data which `PYABSA` was trained on might not accurately reflect the dynamics in my data.
Finetuning the model by Lüdcke et al. resulted in a weighted average accuracy of 86%, with a F1 score of 92% for neutral statements and 30% for negative statements.
Finally, the `Keras`-based neural network generated the best result among the deep learning models:

Label	**Precision**	**Recall**	**F1**	**Support**
Non-negative	0.90	0.95	0.92	2123
Negative	0.37	0.24	0.30	282
Accuracy			0.86	2405
Macro average	0.64	0.60	0.61	2405

Table 4: **Performance of the Keras-based neural network on gold labels:** The Keras-based neural network achieves an accuracy of 86%.

When tested on the gold and silver labels combined, the `Keras`-based neural network achieved a weighted average accuracy of 93%. When tested only on the gold labels, it featured a weighted average accuracy of 86%. Here, precision and recall were more balanced than for the ensemble model.

Overall, the Snorkel labeling functions jointly outperformed the machine learning and deep learning models. Due to the smaller size of the newspaper dataset, the models may not have been able to generalize beyond the diverse set of labeling heuristics. Hence, the weak sentiment labels were selected for further analysis and visualization.

4.1.2 Twitter analysis

Weak labeling: For the Twitter analysis, I established twelve Snorkel labeling functions with an accuracy of over 45%.

LF	**Coverage**	**Conflicts**	**Accuracy**
textblob	1.0	0.80	0.52
hashtags	0.03	0.03	0.91
force	0.01	0.01	0.85
lies	0.02	0.02	0.79
other_neg	0.03	0.03	0.87
punctuation	0.05	0.05	0.69
log_reg_cv_lf	1.0	0.80	0.62
svm_cv_lf	1.0	0.80	0.61
xgbr_cv_lf	1.0	0.80	0.58
rf_cv_lf	1.0	0.80	0.62
xgbr_spacy_lf	1.0	0.80	0.49
bert_mdraw_lf	1.0	0.80	0.57

Table 5: **Performance of the Snorkel labeling functions:** This table shows all labeling functions, including the share of documents they apply to (respectively do not abstain from), the share of other labeling functions they are in conflict with, and their accuracy. A custom-defined function to detect negatively connoted hashtags shows the strongest performance.

For the Twitter dataset, my customized label model achieved the following performance when applied to the manually labelled test set:

Label	**Precision**	**Recall**	**F1**	**Support**
Neutral	0.65	0.78	0.71	486
Positive	0.50	0.11	0.18	36
Negative	0.59	0.47	0.53	346

Accuracy			0.63	868
Macro average	0.58	0.45	0.47	868

Table 6: **Performance of the Snorkel labeling functions:** While the performance for neutral and negative statements is sufficiently strong, the labeling functions frequently fail to correctly identify positive statements, indicated by a low recall.

The silver labels output by the label model featured a weighted average accuracy of 63%. The Snorkel labeling functions performed slightly subpar for the tweets when compared to the newspaper articlesBoth precision and recall were moderately high for neutral and negative statements. For positive statements, however, the recall was significantly lower than the precision. As for the newspaper analysis, the problem was subsequently framed as a binary classification problem.

Machine learning models: Among the standalone machine learning models, the ensemble voting model generated the best result in terms of the F1 macro:

Label	Precision	Recall	F1	Support
Non-negative	0.82	0.83	0.82	1338
Negative	0.72	0.70	0.71	832
Accuracy			0.78	2170
Macro average	0.77	0.76	0.77	2170

Table 7: **Performance of the ensemble model on gold labels:** The ensemble model achieves an accuracy of 78%.

When tested on the gold and silver labels combined, the ensemble model achieved a weighted average accuracy of 91%. When tested only on the gold labels, it featured a weighted average accuracy of 78%.

Deep learning models: Among the four pretrained models, the model by Guhr performed best with a weighted average accuracy of 62% and a F1 score of 72% for neutral statements and 40% for negative statements.
Finetuning the model by Guhr resulted in a weighted average accuracy of 74%, with a F1 score of 78% for neutral statements and 67% for negative statements. Finally, the `Keras`-based neural network generated the best result among the deep learning models:

Label	Precision	Recall	F1	Support
Non-negative	0.81	0.87	0.84	1338

Negative	0.76	0.68	0.72	832
Accuracy			0.80	2170
Macro average	0.79	0.77	0.78	2170

Table 8: **Performance of the Keras-based neural network on gold labels:** The Keras-based neural network achieves an accuracy of 80%.

When tested on the gold and silver labels combined, the `Keras`-based neural network achieved a weighted average accuracy of 88%. When tested only on the gold labels, it featured a weighted average accuracy of 80%.

Overall, the machine learning and deep learning models outperformed the Snorkel labeling functions, likely because they were trained on a significantly larger dataset than for the newspaper analysis. In this case, the deep learning models even slightly improved the classification performance versus the machine learning models.

4.2 Insights

4.2.1 Newspaper analysis

The Covid-19 pandemic has shaped the newspaper coverage from early 2020 until today like no other event within the past decades. In total, German-language newspapers in Switzerland have published nearly 87'000 articles on the subject between January 2020 and April 2022 based on my query of the *SMD* API. As visible on Figure 6, a majority stems from leading national e-papers and traditional newspapers, such as *Cash.ch, Blick, Neue Zürcher Zeitung (NZZ),* and *20 Minuten*.

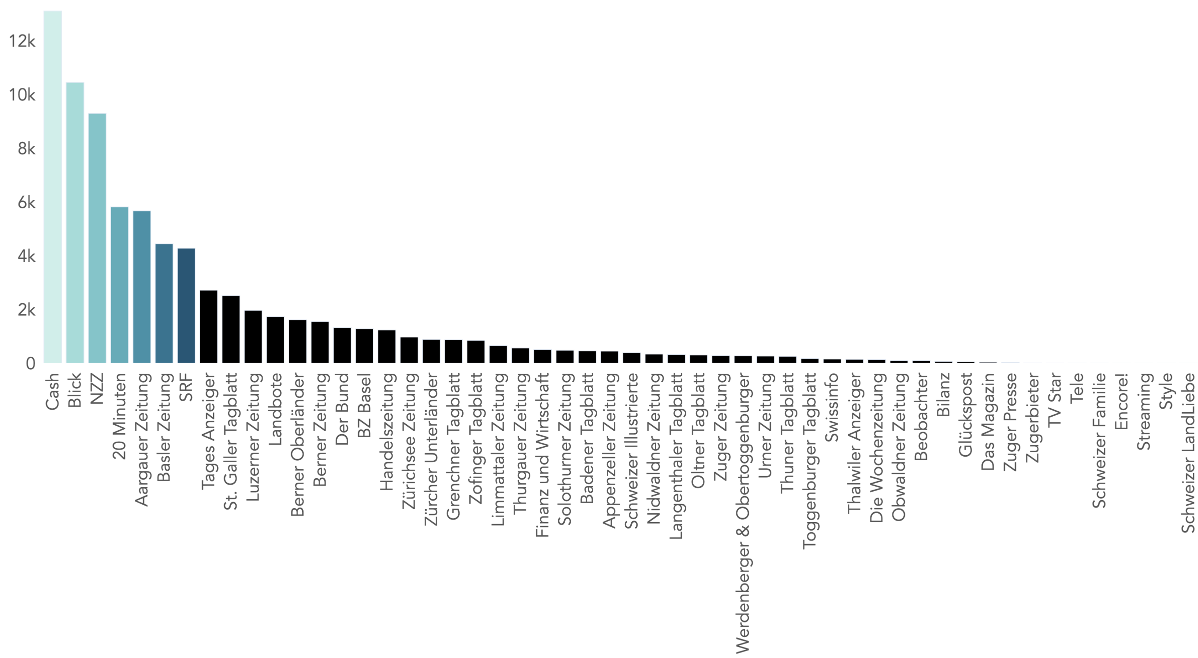

Figure 6: **Total number of Covid-related newspaper articles between January 2020 and April 2022 by newspaper:** Leading Swiss newspapers published more than 5'000 articles directly relating to Covid-19 over the past two years. *Source: SMD API query*

In the context of the pandemic, newspapers have assumed substantial sociopolitical responsibilities by acting as public opinion makers. Newspapers have been criticized on two major fronts, one concerning the volume and the other concerning the content published with regards to the pandemic. In terms of volume, readers have taken issue with the dominance of Covid-related reporting, especially in comparison to other topics [22]. However, when considering the number of Covid-related articles published in relation to the number of hospitalizations, as reported by the Federal Office of Public Health in Switzerland [25], or Google search trends [31] on Figure 7, a disproportion of Covid-related articles cannot necessarily be confirmed.
Based on my query of the *SMD* API, the daily number of Covid-related articles published has undergone fluctuations throughout the past two years. It experienced three peaks, one during the first lockdown in April 2020, a second during November 2020, as public health measures were tightened to obviate a second lockdown, and a third during the Covid-19 act vote in November 2021, which determined the extent to which the Federal Council could maintain necessary measures to manage Covid-19. Generally, the newspaper coverage of the pandemic slowed down during summer months, where incidence counts were relatively low and measures comparably relaxed.

The daily number of hospitalizations due to the pandemic and the Google search trend for the keyword Corona show similar trendlines as the number of Covid-related articles published. A slight decoupling from the hospitalizations can be observed during early 2022. Presumably, the decreased news value of Covid-19 during this time is driven by the unfolding of the Russia-Ukraine crisis, which largely consumed the interest of the media and the general public.

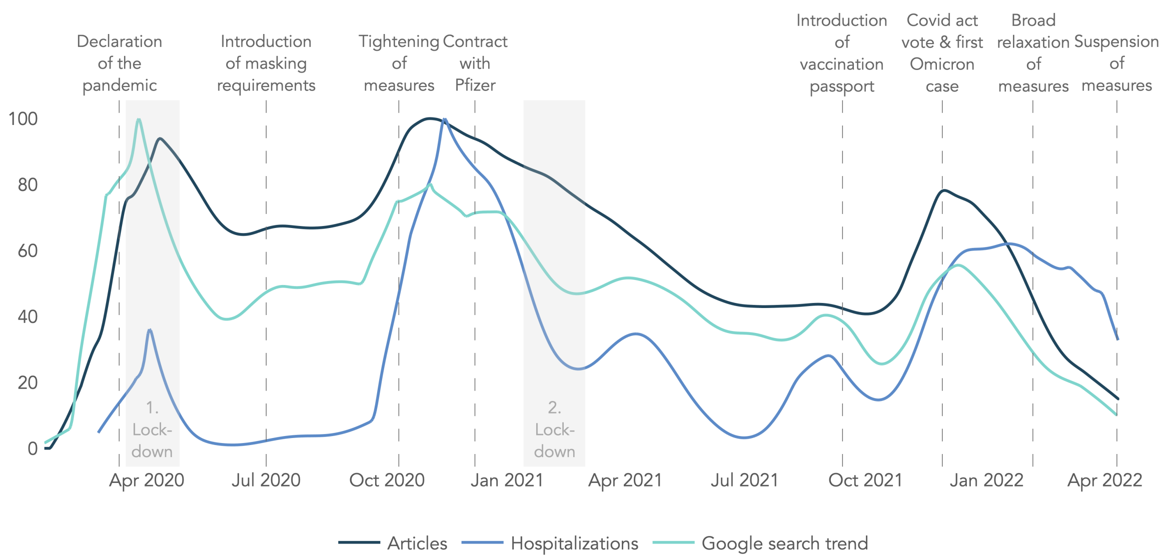

Figure 7: **Daily number of Covid-related newspaper articles between January 2020 and April 2022 compared to hospitalizations and Google search trends for the keyword *Corona*, each indexed to 100:** The media relevance of Covid-19 follows a similar trendline as the number of Covid-related hospitalizations and Covid-related Google search trends. *Source: SMD API query, timeseries analysis*

Newspapers have not only experienced criticism pertaining to the volume of Covid-related articles, but also pertaining to the article content, in particular to what extent articles have provided an objective account of Covid-19. In this context, it is worth investigating which political agents the newspaper coverage has focused on and to what extent these agents have been viewed in a positive or negative light.
Based on my entity recognition approach, Figure 8 shows which agents the newspaper coverage has concentrated on. It is unsurprising that key decision makers throughout the pandemic, such as Switzerland's *Bundesrat* – the Federal Council – and Alain Berset – its Minister of Health - are mentioned most frequently. When considering other officials, it is interesting to note that Daniel Koch – the former Head of Communicable Diseases in the Federal Office of Public Health – has a strong media presence, despite his retirement at the end of May 2020. This is backed by the suspicions of critics, according to which Koch stepped out

of the spotlight only begrudgingly, indicated by his consultancy work, a book project, and continued public appearances even after his departure from the office [63].

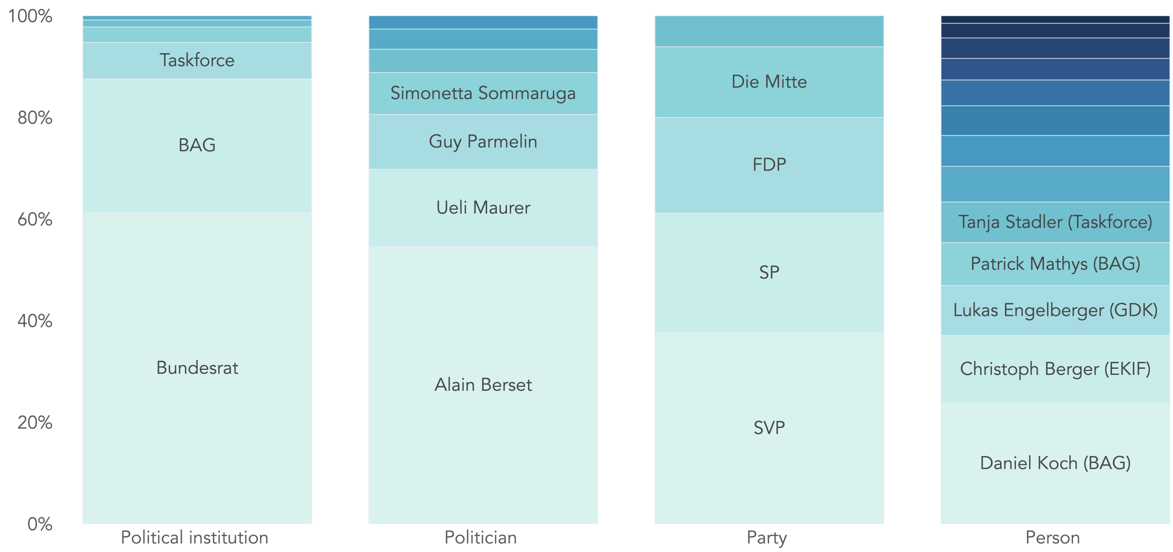

Figure 8: **Share of mentions by group and agent in newspaper articles:** The *Bundesrat* and Alain Berset are mentioned most frequently in newspaper articles. *Source: Entity recognition*

The extent to which key political agents have been considered positively or negatively can be quantified based on my sentiment analysis[2]. Here, I focus only on the most important politicians, institutions, and officials for the sake of brevity. As visible on Figure 9, Ueli Maurer and Alain Berset are mentioned most negatively. This is unsurprising given their central positions as members of the Federal Council and their roles as Minister of Finances and Minister of Health respectively. Overall, when considering the sentiment for all seven of the Federal Council members, the ranking closely tracks the sympathy ranking as per a survey by the *SRG* which was conducted in October 2021 [69]: Federal Council members with lower sympathy rankings tend to be more strongly criticized during Covid-19. The only exception here is Alain Berset, who, despite high sympathy ratings, has been heavily criticized throughout the pandemic.
Figure 9 also shows that criticisms are largely attributed to either members of the Federal Council or institutions as a whole, rather than individuals holding roles within these institutions: Federal Council members like Alain Berset and Ueli Maurer as well as institutions like the *Taskforce* show the strongest negative sentiment.

[2]As described in Chapter 3.2.3, the sentiment analysis problem is framed as a binary classification problem, distinguishing between negative and non-negative statements only.

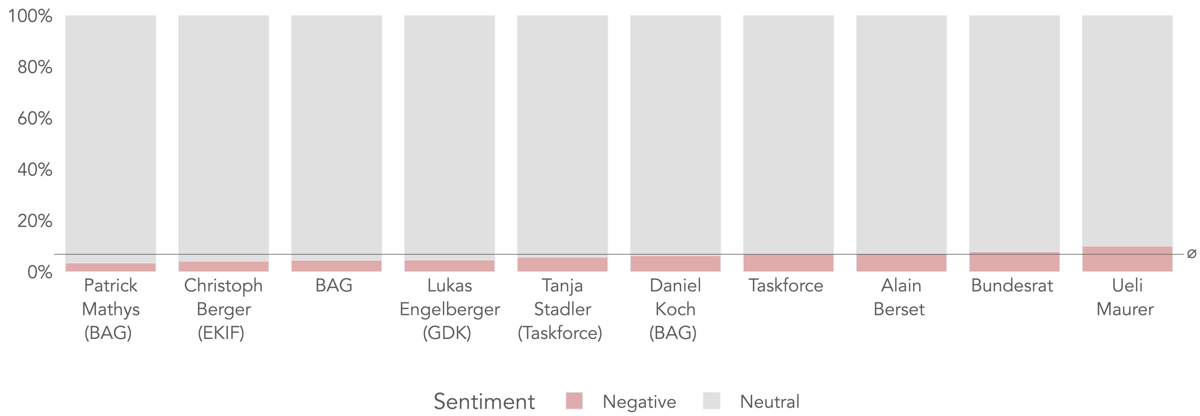

Figure 9: **Sentiment split by agent:** Newspaper articles negatively mention political agents in less than 20% of the instances. The weighted average share of negative mentions across all agents, indicated by ∅, lies at approximately 7%. *Source: Sentiment analysis*

My sentiment analysis for each political agent can be further broken down by publisher, as can be seen on Figure 10. Here, it is interesting to note that both the *SRG* and *Ringier*, on average, level less criticism against Alain Berset than other publishers. Both the *SRG* and *Ringier* have been exposed to charges regarding their pro-government coverage of Covid-19. While the *SRG* is independent from the Swiss government in terms of its legal form and ownership structure, it is connected to the government through its mandate, board, and funding [13].

Similarly, *Ringier* has been criticized for pro-government reporting, after a video of a conference leaked, where the CEO Marc Walder called upon his journalists to support the government throughout Covid-19 in their coverage [71].

As Alain Berset can be considered the main representative of the government during the pandemic in his role as Minister of Health, my sentiment analysis serves some support to the charges drawing into doubt the independence of Swiss media.

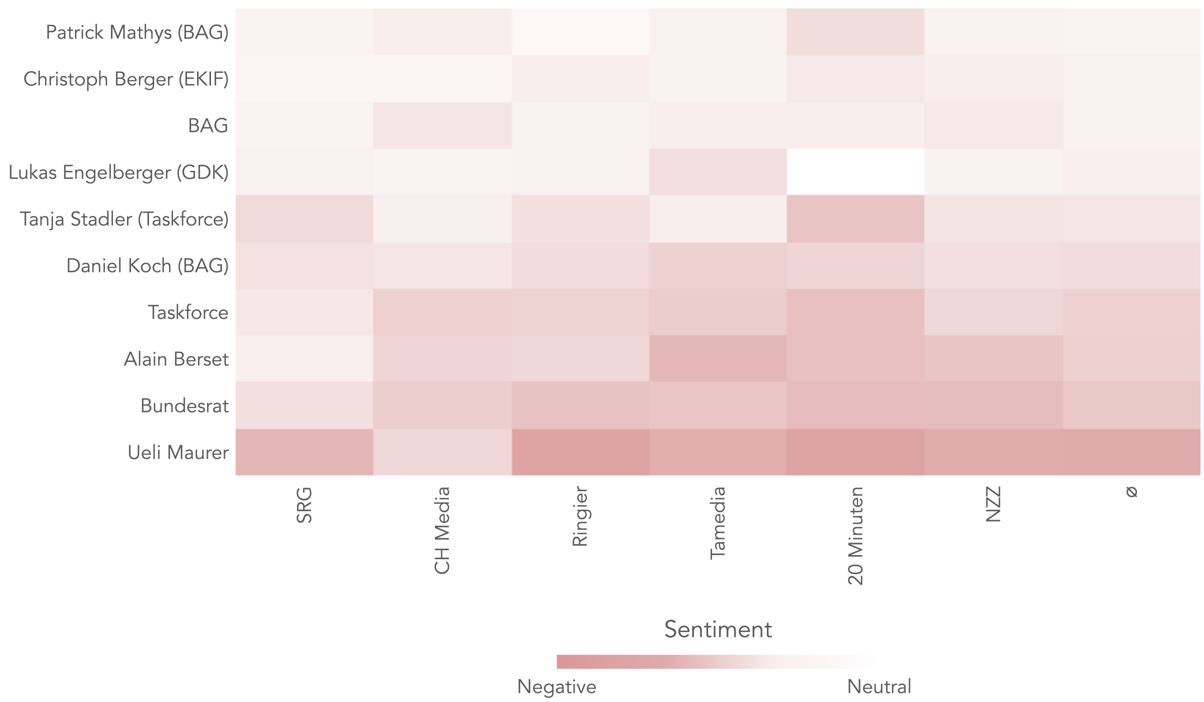

Figure 10: **Average sentiment by publisher on the x-axis and agent on the y-axis:** The weighted average sentiment across publishers, indicated by ∅, is lowest for Ueli Maurer and highest for Patrick Mathys. *Source: Sentiment analysis*

The sentiment for each political agent can also be shown over time based on my timeseries analysis, which relies on a Box1DKernel and LOWESS. Here, I focus only on the most important politicians and institutions given sample size considerations. For each agent, I give an interpretation of the sentiment curve over time based on the outputs of my topic modeling analysis, which relies on NMF.

For most agents, a seasonality can be observed. During summer months, when the article volume related to Covid-19 tends to decrease due to a more relaxed public health situation, the sentiment tends to increase.

Let us first consider the timeseries for key politicians on Figure 11, namely Ueli Maurer and Alain Berset.

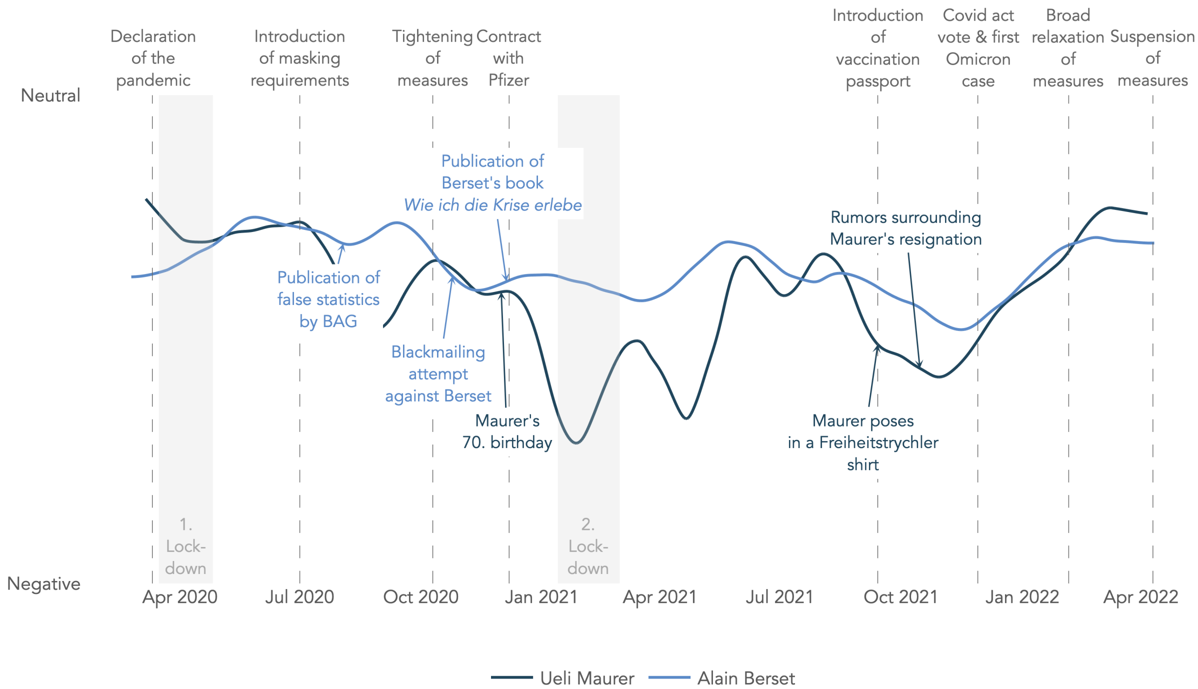

Figure 11: **Average sentiment between January 2020 and April 2022 for Alain Berset and Ueli Maurer:** The sentiment curve for Ueli Maurer is nearly consistently below the sentiment curve for Alain Berset. *Source: Sentiment analysis, timeseries anlaysis, topic modeling*

Particularly at the outset of the pandemic, Alain Berset – the Minister of Health – was celebrated as man of the moment by the media, driven by his strong leadership presence.
However, various slip-ups in Berset's *BAG* - the Federal Office of Public Health – soon gave rise to criticism concerning a lack of digitalization, chaotic data management, and confusing guidelines, indicated by a dropping sentiment curve.
Towards the end of 2020, charges against Berset largely targeted Switzerland's idiosyncratic strategy in dealing with the pandemic, which was characterized by hesitancy and a late second lockdown when compared to neighboring European countries.
After the second lockdown, Berset was criticized for selective issues, including the slow vaccination progress, an uncoordinated division of tasks between the federation and the cantons, and a failure to consider economic interests when determining public health measures.
Towards the end of the pandemic in 2022, opinions on Berset's performance were divided, both with regards to the past two years and with regards to the comparably rapid relaxation of measures. Hence, the sentiment curve reaches level slightly below the pre-Covid level. Overall, Berset was the face of the

pandemic in Switzerland par excellence. His attempts to find a middle ground between the right-wing and left-wing fraction were criticized from both sides. While right-wing fractions considered him a dictator with a disproportional focus on public health issues, left-wing parties considered him a passive-reactive policy maker. Despite these charges, the criticism by and large seems to have left Berset untouched, presumably due to his skillful media presence. Even after Covid-19, Berset continues to outperform his fellow Federal Council members in the sympathy rankings [69].

Just as Berset, Ueli Maurer – the Minister of Finances – was praised by the media at the beginning of the pandemic due to his fast, unbureaucratic reaction to the economic crisis.
However, soon Maurer became a talking point through his wayward stances compared to fellow Federal Council members, reflected by a dropping sentiment curve. For example, Maurer declared that he does not understand and would not use the federal SwissCovid app, which was introduced for digital contact tracing purposes. He also opposed a second lockdown as the only member of the Federal Council.
After the second lockdown, the newspaper coverage of Maurer was mixed. On the one hand, his portrayal of the deficits in the Federal Treasury were criticized as exaggerated. On the other hand, Maurer found support from the trade sector due to his focus on economic interests.
The sentiment curve for Maurer reaches another trough in September 2021, when he posed in a Freiheitstrychler shirt and therefore provoked associations with Covid skeptics as well as retirement rumors. Many critics interpreted this scene as a violation of the Swiss Kollegialitätsprinzip, according to which the seven members of the Federal Council must jointly reach decisions behind closed doors and unanimously represent these decisions in front of the general public.
Towards the end of the pandemic in 2022, Maurer was only seldomly mentioned in the context of Covid-19. In sum, he proved to be an agile, unbureaucratic Minister of Finances and a maverick within the Federal Council. This secured him the sympathies of the population rejecting strict public health measures in favor of economic interests as well as Covid skeptics. Despite several fluctuations throughout the past two years, the sentiment curve for Maurer finally settles at the pre-Covid level.

Let us now consider key institutions on Figure 12. The sentiment curve for institutions is less volatile than for individuals.

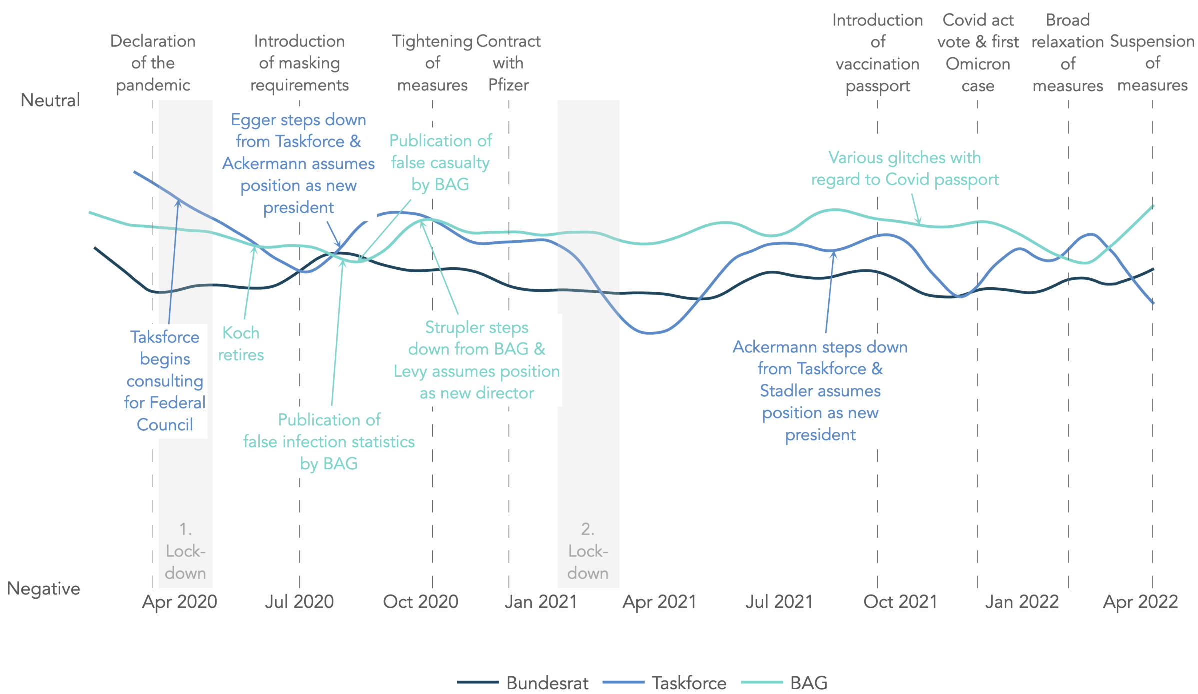

Figure 12: **Average sentiment between January 2020 and April 2022 for key institutions:** The sentiment curves for institutions are generally less volatile than the sentiment curves for individuals. *Source: Sentiment analysis, timeseries analysis, topic modeling*

For the *Bundesrat* – the Federal Council –, the sentiment curve roughly follows the phases of the pandemic. While the trust placed into the *Bundesrat* was great at the outset of the pandemic, later phases of the pandemic unearthed fundamental issues, which the media picked up on. For example, towards the end of 2020, many critics questioned the broad decision-making authority awarded to the *Bundesrat* and the *BAG* under emergency law and demanded a more diversified structure of governmental organs.
After the second lockdown, other topics came into focus, such as the division of tasks between the federation and the cantons, the balance between public health, economic, and social interests in decision-making throughout the pandemic, and the social divide in wake of the Covid-19 act.
Overall, Switzerland has come through the pandemic with a strong economic position and a mediocre public health position [41], which is reflected in that the sentiment curve for the *Bundesrat* does not fully recover to the pre-Covid level.

At the beginning of the pandemic, the *Bundesamt für Gesundheit (BAG)* – the Federal Office of Public Health – was relatively rarely mentioned by the

media. However, various slip-ups soon gave rise to criticism, concerning a lack of digitalization, chaotic data management, confusing guidelines, and a lack of expert knowledge under its director Pascal Strupler.
With the resignation of Strupler and the start of Anne Levy in October 2020, the sentiment curve stabilizes. After the second lockdown, the charges against the *BAG* largely focused on the slow vaccination progress and the *BAG's* hesitancy to tighten measures throughout the spread of Omicron.
Towards the end of the pandemic in 2022, reviews for the *BAG* were largely negative, indicated by a drop in sentiment around the time it was announced that Switzerland would lift all measures by April 2022. In sum, it was criticized for its lack of strategic preparation for a pandemic, its slow decision making, and a lack of expert knowledge.

The sentiment curve for the *Taskforce* – a gremium instituted at the end of March 2020 to serve as an advisory for public authorities, especially the *BAG* – initially features a similar progression as the curve for the *BAG* itself. However, soon concerns emerged regarding the disagreement amongst *Taskforce* members as well as between the *Taskforce* and the *BAG*. As for the *BAG*, a change of leadership provided some stabilization with Martin Ackermann taking office in August 2020.
However, after the second lockdown, the sentiment curve for the *Taskforce* dips below the curve for the *BAG* and does not recover. Points of criticism from the media included continued public disagreements with the *BAG*, alarmist prognoses, and a unilateral consideration of public health interests. In particular, it was never fully clarified whether the *Taskforce* was to serve the *BAG* as an internal advisory or whether they could independently make media-effective statements. These charges culminated in the discussion of a muzzle for the *Taskforce* [72], which caused some restraint among *Taskforce* members throughout the second trimester of 2021, during which the sentiment curve recovers slightly. With the advent of Omicron, the *Taskforce* rejoined the conversation, however, again with false prognoses. Overall, the *Taskforce's* record appears mainly negative. This is driven by a lack of understanding of the precise role of the *Taskforce*, substantial disagreements with the *BAG*, and overshooting prognoses.

4.2.2 Twitter analysis

When the *WHO* was first informed regarding Covid-19 at the end of December 2019, the Swiss Twittersphere took no notice. Based on my query of the Twitter API, the topic began trending only when the situation began to escalate in Italy in February 2020.
When considering the number of Covid-related tweets which mention selected political agents in relation to the number of hospitalizations [25] or Google search trends [31] on Figure 13, interesting dynamics emerge. For example, while the Google search trend peaked during the first lockdown, the number of daily tweets and hospitalizations did not reach their highest levels at this point in time. Similarly, while hospitalizations hit an all-time high shortly before the

second lockdown, the Google search trend and the tweets followed at lower levels. What stirred the Twitter users the most was the Covid-19 act in November 2021.
Figure 13 also highlights the fact that the behavior of the general online audience, indicated by the Google search trend, does not necessarily reflect the behavior of Twitter users.

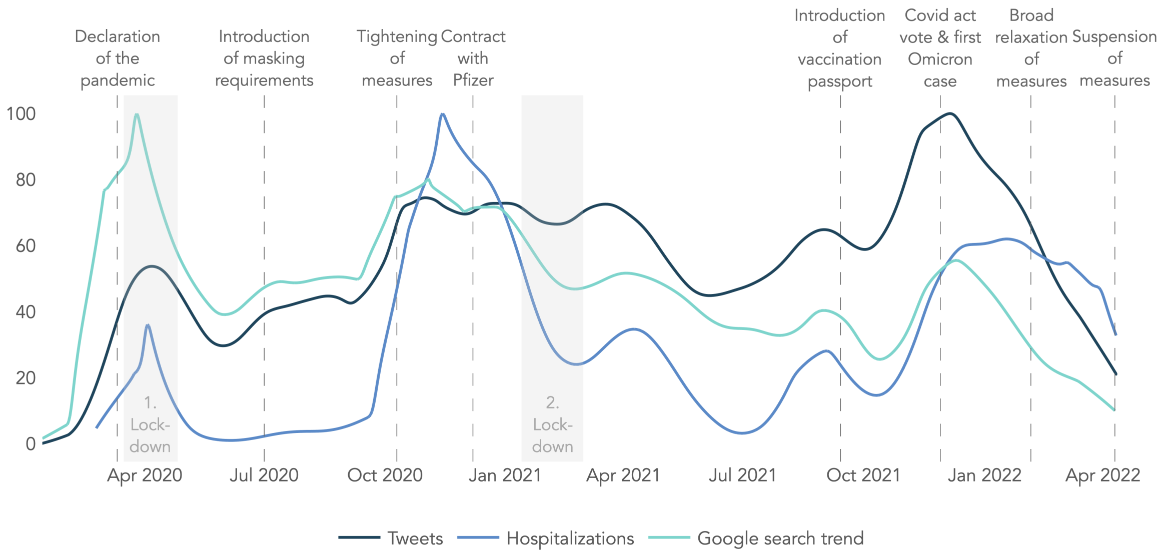

Figure 13: **Daily number of Covid-related tweets mentioning selected agents between January 2020 and April 2022 compared to hospitalizations and Google search trends for the keyword *Corona*, each indexed to 100:** The relevance of Covid-19 for the Twittersphere follows a slightly different curve than the number of Covid-related hospitalizations and Covid-related Google search trends. *Source: Twitter API query, timeseries analysis*

While Twitter cannot be considered representative of the general public [52], it nonetheless has been recognized as a primary online avenue for information gathering, opinion formation, and persuasion [42] [26]. In the context of the pandemic in Switzerland, it is worth investigating which political agents the Twittersphere has targeted and to what extent these agents have been viewed positively or negatively.
Covid-related tweets show a more extreme focus on key decision makers than newspapers, as can be seen on Figure 14. The most frequently mentioned agents include Switzerland's *BAG*, the *Bundesrat*, and Alain Berset. When considering other officials, it is interesting to note that Daniel Koch again has a strong presence on Twitter. The same goes for Tanja Stadler – the head of the *Taskforce* – who took office in August 2021. This suggests that, despite their relatively short terms of office, these two agents were perceived as the public faces of their

respective institutions.

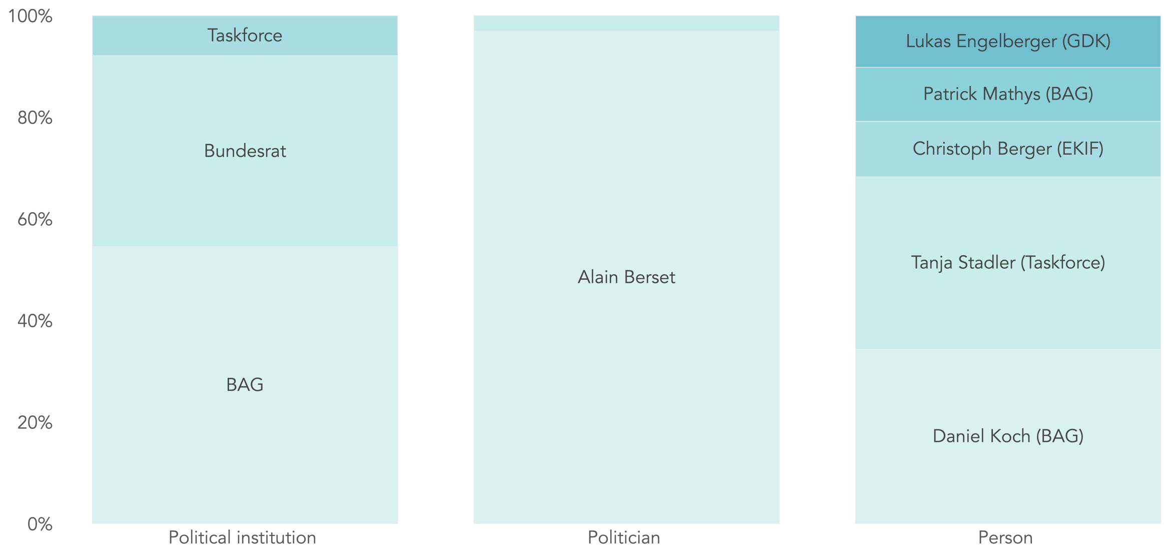

Figure 14: **Share of mentions by group and agent in tweets:** The *BAG*, the *Bundesrat*, and Alain Berset are mentioned most frequently in tweets. *Source: Twitter API query*

The extent to which key political agents have been considered in a positive or negative light is visible on Figure 15. Ueli Maurer, Christoph Berger, and Alain Berset are mentioned most negatively. Berset clearly occupies the center stage in this context, with a total number of nearly 40'000 mentions. Maurer and Berger – the president of the Federal Commission for Vaccination Issues – find themselves at the heart of a smaller-scale, yet equally fierce criticism movement.

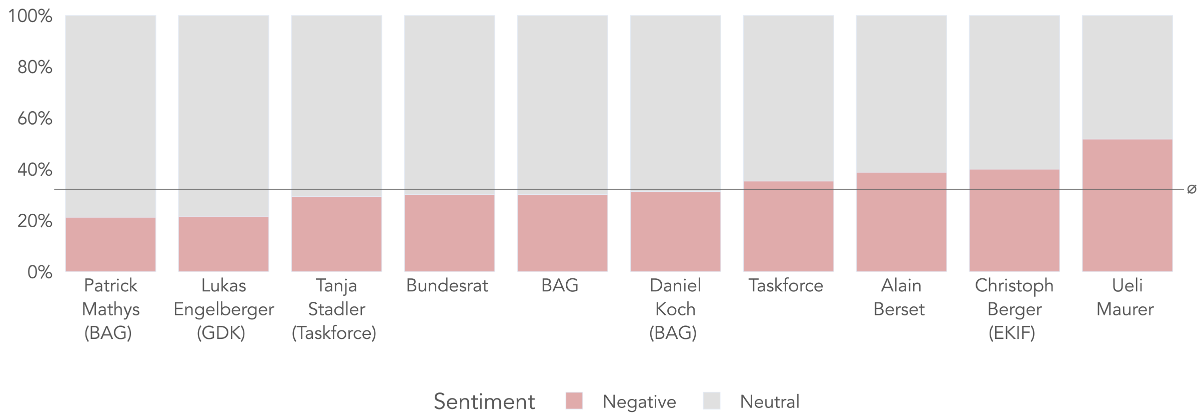

Figure 15: **Sentiment split by agent:** Depending on the political agent, tweets express criticism in between 20% and 55% of the instances. The weighted average share of negative mentions across all agents, indicated by $\varnothing$, lies at approximately 33%. *Source: Sentiment analysis*

Let us now consider the sentiment for each political agent alongside key points of criticism. The sentiment curve for key politicians, namely Ueli Maurer and Alain Berset, as can be seen on Figure 16.

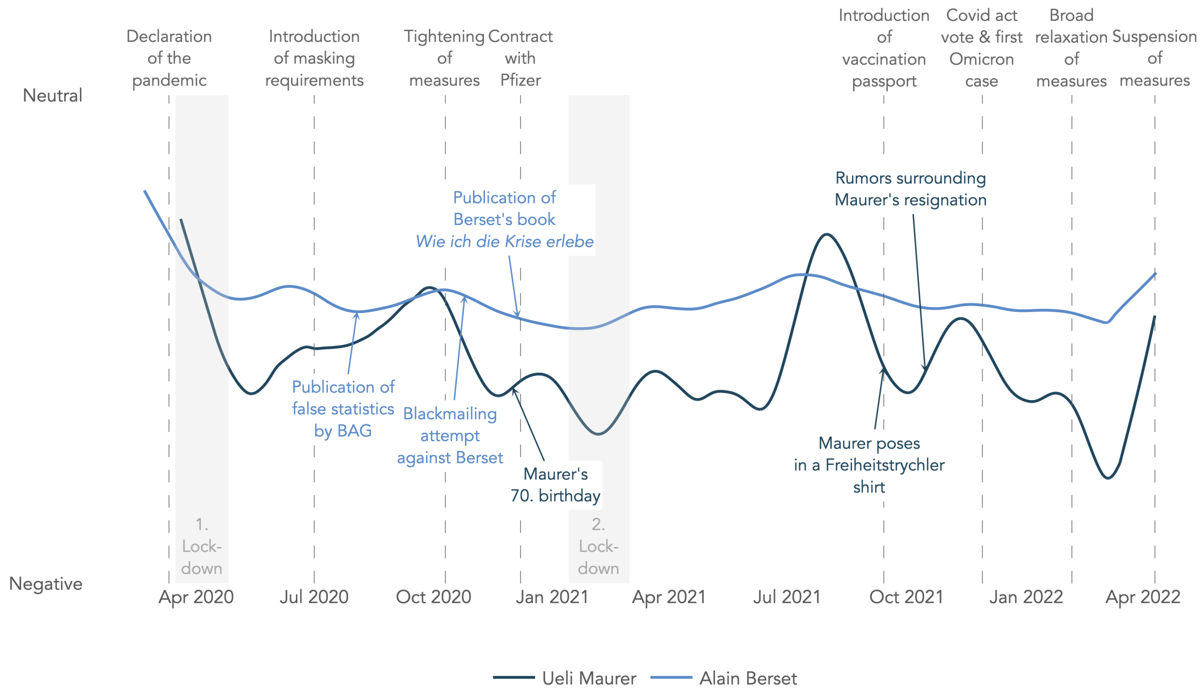

Figure 16: **Average sentiment between January 2020 and April 2022 for Alain Berset and Ueli Maurer:** The sentiment curve for Ueli Maurer is nearly consistently below the sentiment curve for Alain Berset. *Source: Sentiment analysis, timeseries anlaysis, topic modeling*

The sentiment curve for Alain Berset – the Minister of Health – as per the Twitter analysis follows a similar trend to that of the newspaper analysis. Nonetheless, the charges on Twitter were more ambivalent: While opponents of public health measures and the vaccination reproached the excessive restrictions and the two-tier society introduced through the vaccination passport, proponents of public health measures and the vaccination accused Berset of a deficient strategy and a lack of support for vulnerable groups, such as children. Even towards the end of the pandemic in 2022, opinions on Berset's performance remained divided, reflected in the sentiment curve reaching level slightly below the pre-Covid level.

For Ueli Maurer – the Minister of Finances –, the sentiment curve as per the Twitter analysis also shows a similar behavior to that of the newspaper analysis. Points of criticism largely overlapped with the newspaper coverage, focusing on Maurer's wayward actions which were frequently interpreted as violations of the Swiss Kollegialitätsprinzip. Nonetheless, Maurer also found some support by opponents of public health measures and the vaccination. In sum, Maurer's record was mixed, mirrored in the sentiment curve settling slightly below the pre-Covid level.

Let us now consider key institutions, as visible on Figure 17.

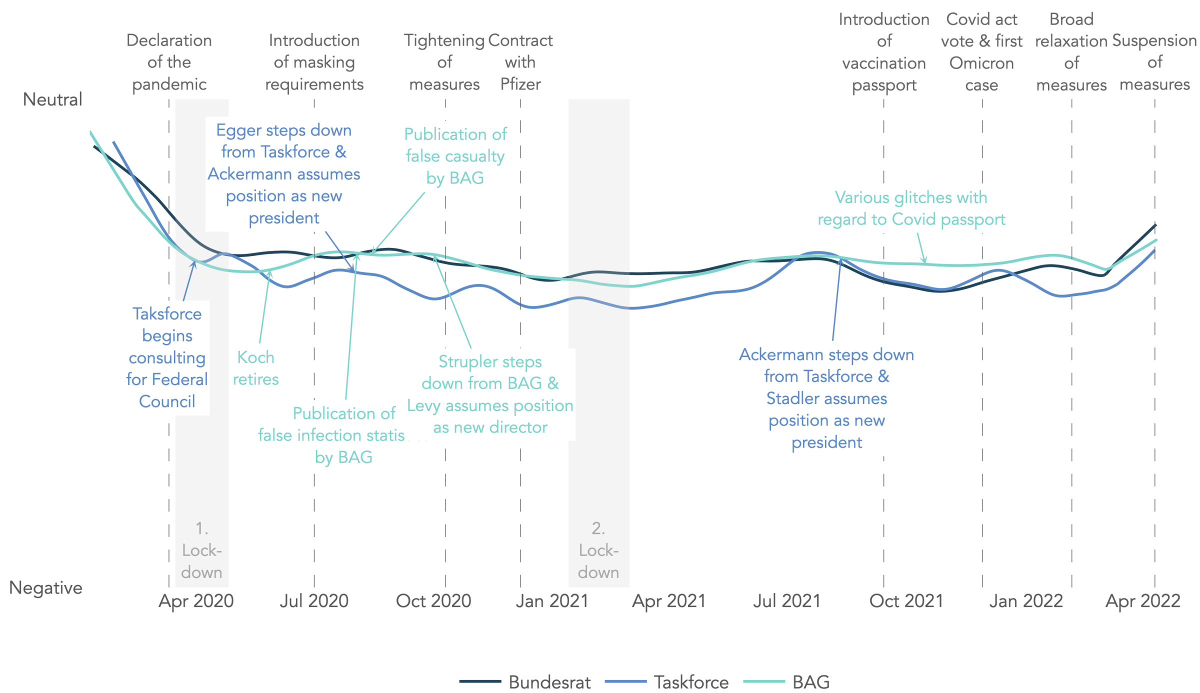

Figure 17: **Average sentiment between January 2020 and April 2022 for key institutions:** The sentiment curves for institutions are generally less volatile than the sentiment curves for individuals. *Source: Sentiment analysis, timeseries anlaysis, topic modeling*

For the *Bundesrat* – the Federal Council –, the sentiment curve again roughly follows the phases of the pandemic. In summer 2020, the inconsistent statements surrounding the efficacy of masks made by public authorities formed a trending topic. Towards the end of 2020, opinions toward the *Bundesrat* were split among proponents of federal public health measures and opponents. The social divide became particularly pronounced in fall 2021 in wake of the vaccination passport and the Covid-19 act. Overall, the sentiment curve for the *Bundesrat* does not fully recover to the pre-Covid level.

Points of criticism towards the *Bundesamt für Gesundheit (BAG)* – the Federal Office of Public Health – largely coincided with the newspaper coverage and concern a lack of strategic preparation for a pandemic, slow decision-making, and a lack of expert knowledge. Towards the end of the pandemic in 2022, the sentiment curve for the *BAG* does not completely return to the pre-Covid level.

The sentiment curve for the *Taskforce* again initially features a similar progression as the curve for the *BAG* itself. However, on Twitter, the sentiment curve for the *Taskforce* dips below the *BAG* sooner than in the media reporting. Points of criticism largely overlapped with the newspaper coverage. Overall, the *Taskforce's* record appears mainly negative, reflected in the drop in sentiment around the time it was announced that Switzerland would lift all measures by April 2022.

4.2.3 Comparative analysis

Both newspapers and Twitter represent key platforms that reflect and shape public opinion. How do these two sources compare in terms of focus and sentiment towards political agents throughout the pandemic in Switzerland?
When considering the agents at the center stage of newspaper articles and tweets on Figure 18, first differences emerge. In terms of key politicians, newspaper articles offer a slightly more even coverage across the Federal Council members, whereas tweets are extremely focused on Alain Berset. This suggests that, while many actors pulled strings throughout the pandemic, Berset was primarily held accountable by the general public.

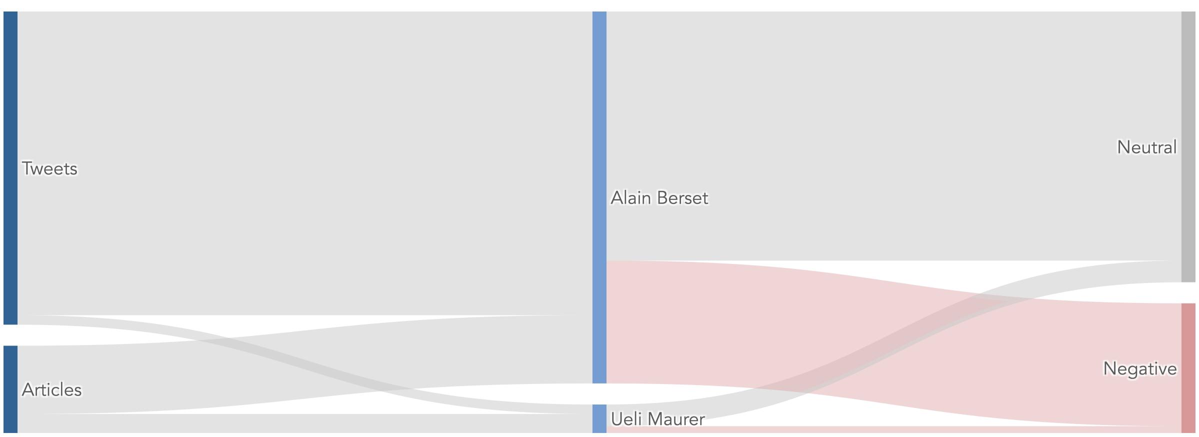

Figure 18: **Number of mentions of key politicians by source and corresponding sentiment:** While tweets strongly focus on Alain Berset, newspaper articles show a more balanced reporting between key politicians. *Source: Entity recognition, sentiment analysis*

With regard to institutions, it is noticeable that the *Taskforce* is considered a secondary matter for both newspaper articles and tweets, as can be seen on Figure 19. Whereas tweets focus on the *Bundesrat* and the *BAG* approximately evenly, newspaper articles clearly focus on the *Bundesrat* to a greater extent.

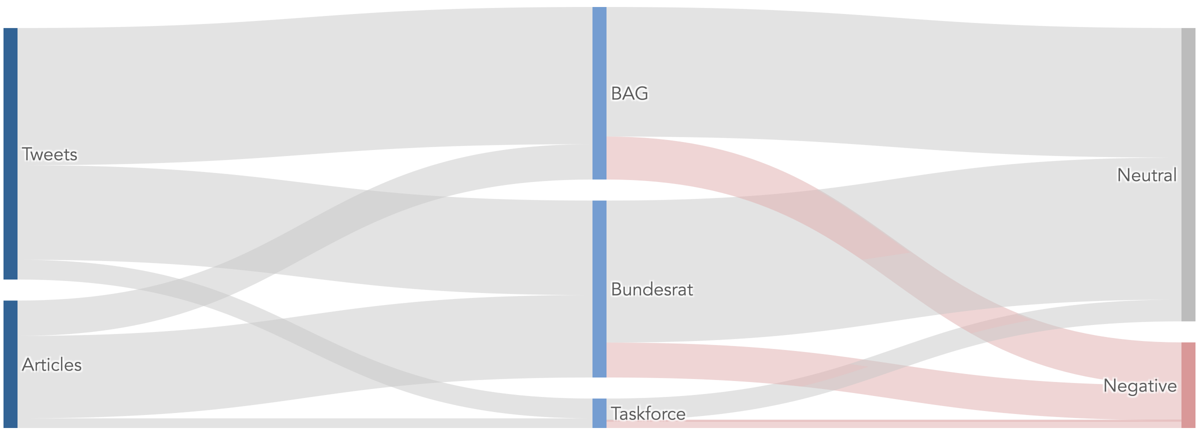

Figure 19: **Number of mentions of institutions by source and corresponding sentiment:** While tweets evenly focus on the *BAG* and the *Bundesrat*, newspapers predominantly refer to the *Bundesrat* in their coverage. *Source: Entity recognition, sentiment analysis*

Lastly, newspaper articles provide a relatively even coverage across different officials, as visible on Figure 20. On Twitter, in contrast, the debate heavily focuses on Daniel Koch and Tanja Stadler.

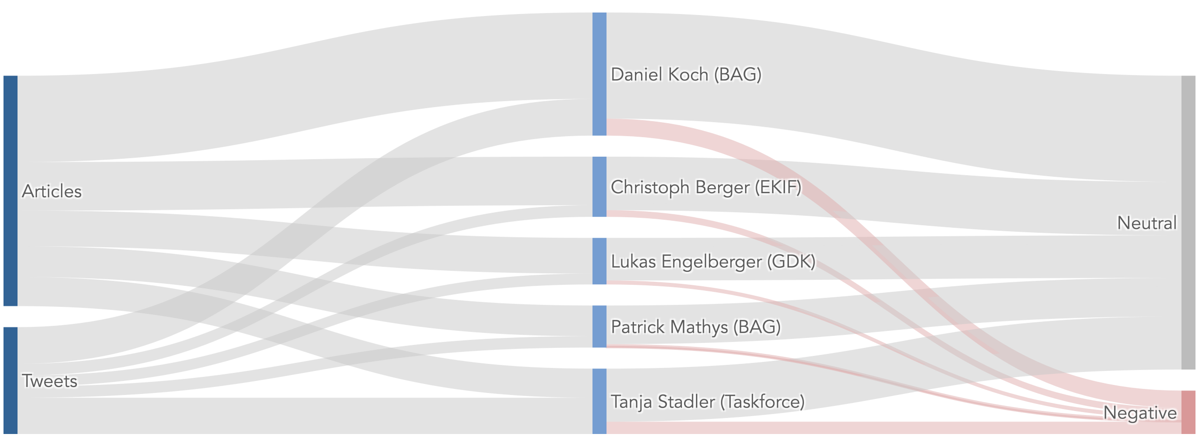

Figure 20: **Number of mentions of officials by source and corresponding sentiment:** While tweets strongly focus on Daniel Koch and Tanja Stadler, newspaper articles show a more balanced reporting between key officials. *Source: Entity recognition, sentiment analysis*

Further differences appear in view of the sentiment expressed towards political agents in newspaper articles and tweets, as can be seen on Figure 21. Based on my sentiment analysis, it becomes apparent that the reporting in Swiss newspapers, by and large, is neutral and homogeneous. In contrast, the agent mentions in tweets are strongly sentiment-laden.

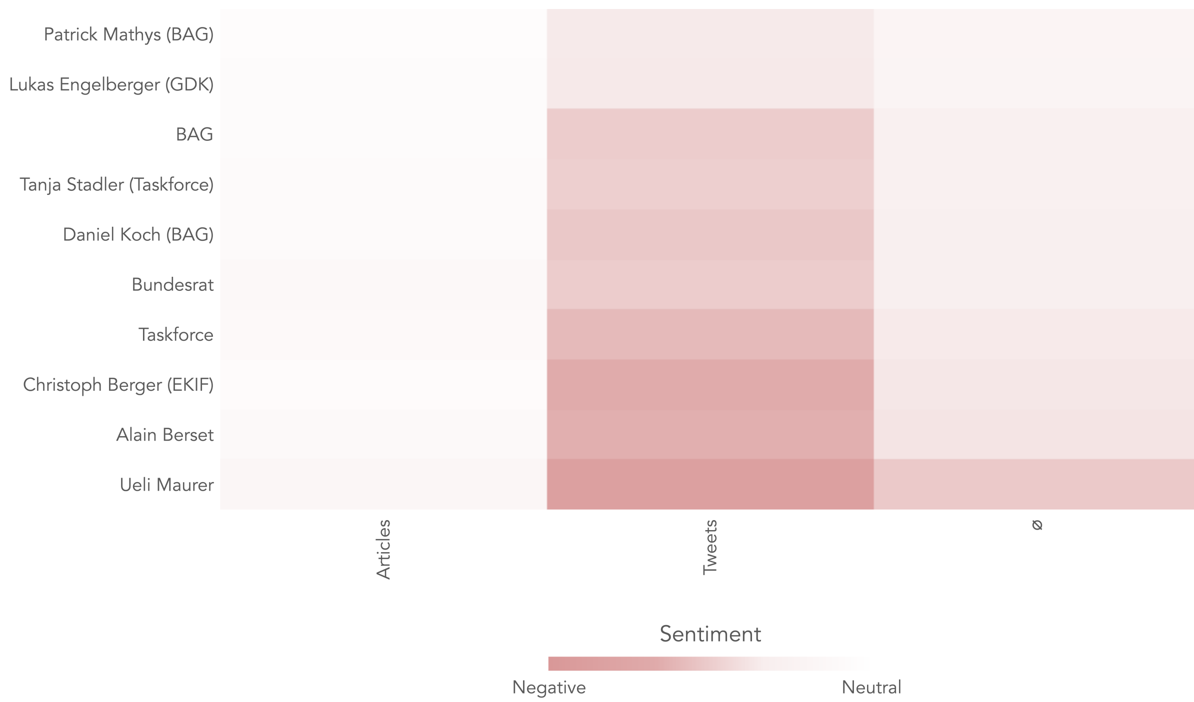

Figure 21: **Average sentiment by source on the x-axis and agent on the y-axis:** While the Twittersphere shows a polarized reaction to several political agents, the newspaper coverage is largely neutral and homogeneous. The weighted average sentiment across both sources, indicated by ∅, is lowest for Ueli Maurer and highest for Patrick Mathys. *Source: Sentiment analysis*

When considering the sentiment rankings for politicians, institutions, and officials side by side, the variations become more apparent. The key politicians of the Covid-19 crisis, Alain Berset and Ueli Maurer, are ranked similarly by newspapers and Twitter, as visible on Figure 22. In both sources, the ratio between the share of negative mentions for Berset and the share of negative mentions for Maurer lies at 1.3 to 1.4.

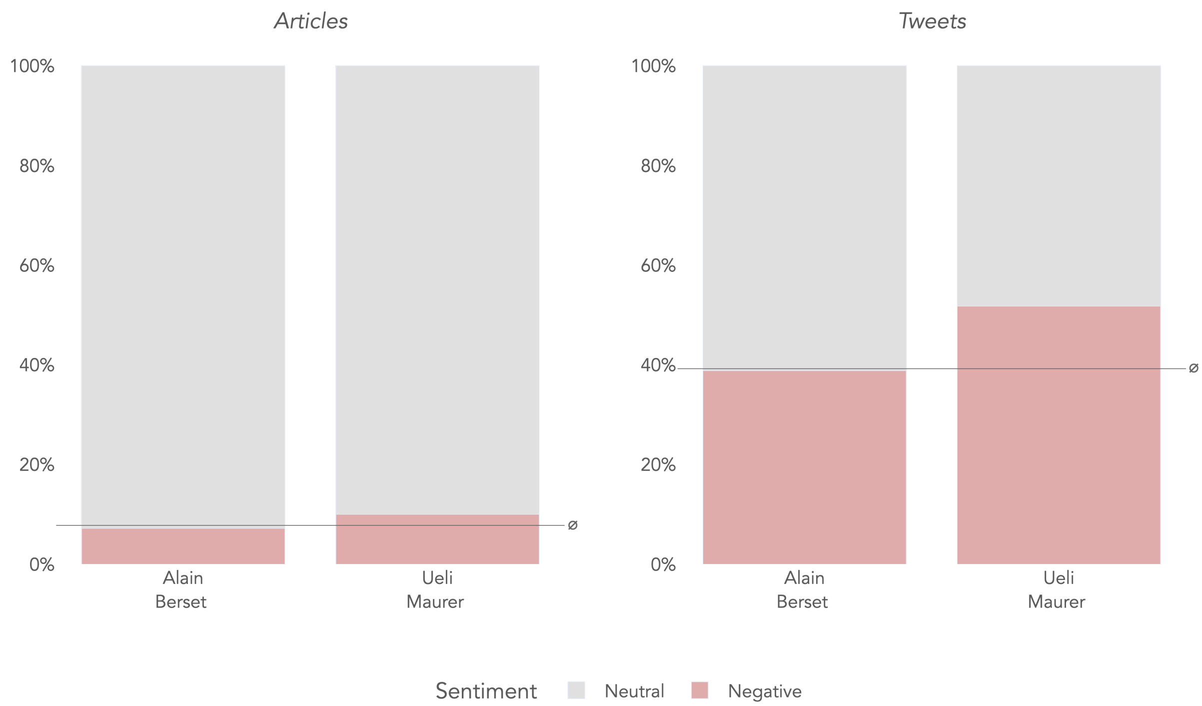

Figure 22: **Sentiment split by source and politician:** While tweets tend to express more criticism, the sentiment ranking between the key politicians is similar across Twitter and newspapers. *Source: Sentiment analysis*

Public institutions receive dissimilar criticism from newspapers and Twitter, as can be seen on Figure 23. While newspapers largely criticize the *Bundesrat* and the *BAG*, Twitter more frequently criticizes the *Taskforce*.

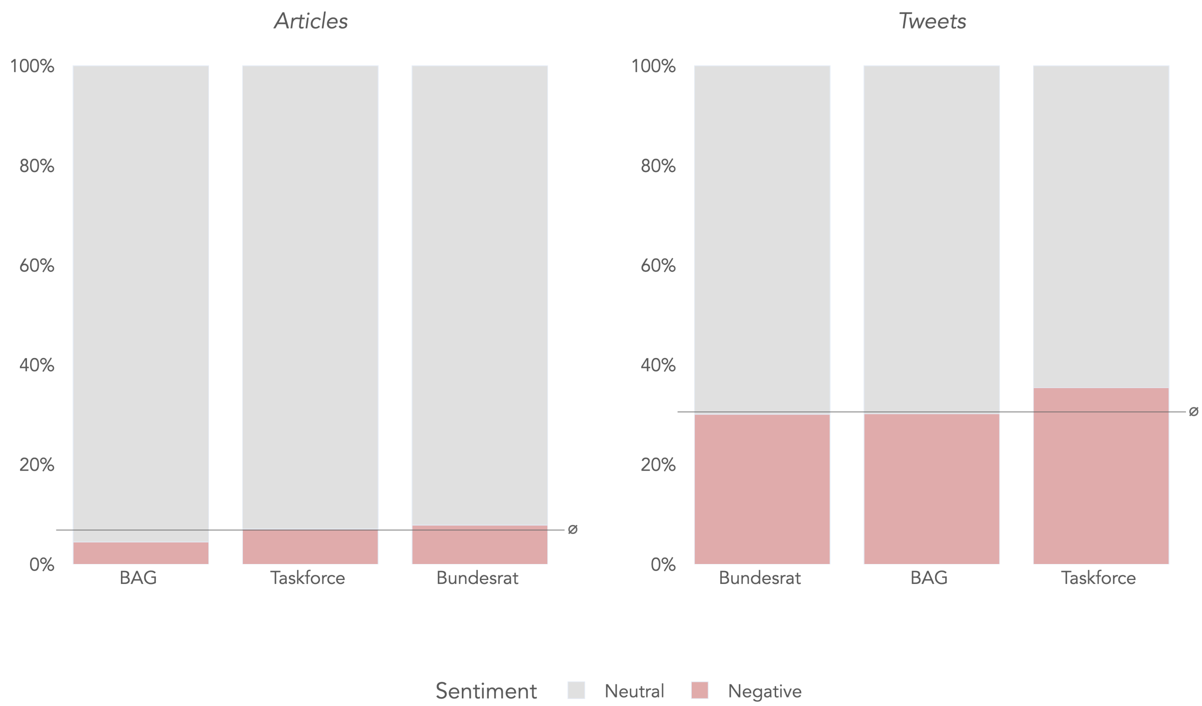

Figure 23: **Sentiment split by source and institution:** While tweets criticize the *Bundesrat* the least among key institutions, it is the most heavily criticized institution in newspaper articles. *Source: Sentiment analysis*

Finally, officials partially receive diverging criticism from newspapers and Twitter, as visible on Figure 24. For newspapers, Daniel Koch and Tanja Stadler are key targets of criticism. On Twitter, this holds true as well. However, in addition, Christoph Berger is a leading recipient of criticism. In particular, issues such as the vaccination for kids, the side effects and efficacy of the vaccine, and the vaccination passport were held against him, based on my topic modeling analysis.

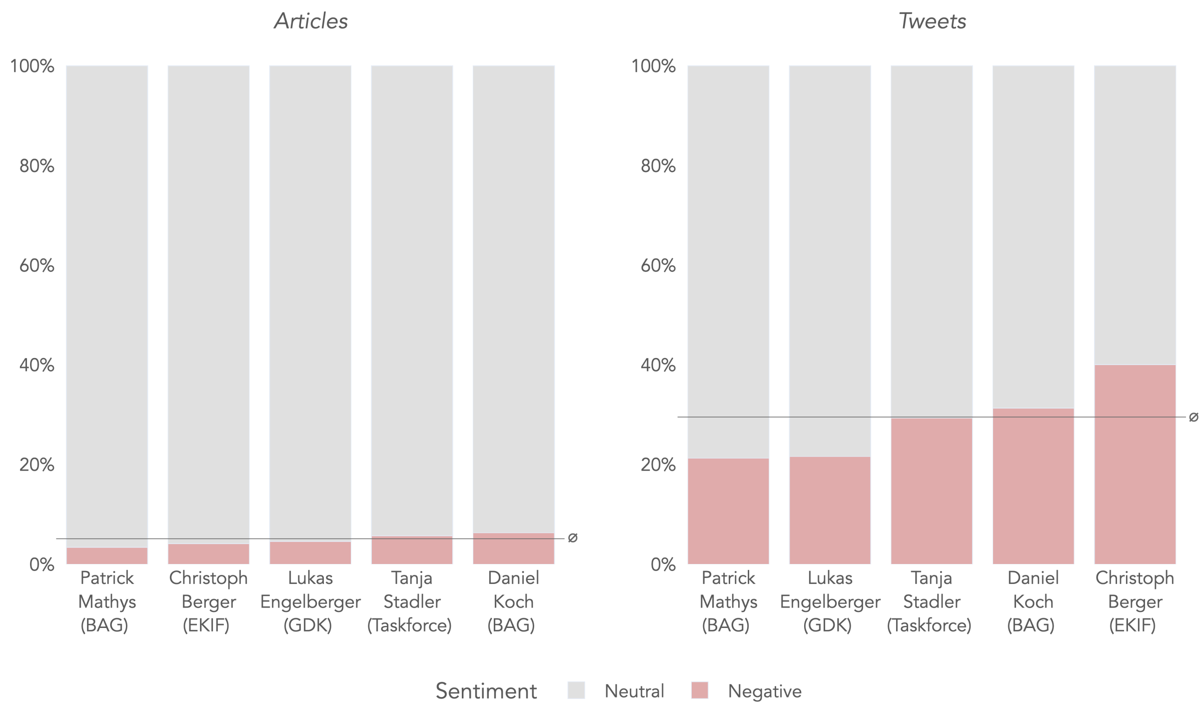

Figure 24: **Sentiment split by source and official:** While tweets criticize Christoph Berger the most among key officials, he is rarely criticized in newspaper articles. *Source: Sentiment analysis*

4.3 Applications

My work provides a value add for society and the company I collaborated with, namely the *Neue Zürcher Zeitung (NZZ)*.

In terms of sociopolitical contributions, this project firstly combats misinformation amongst citizens that consume newspapers and other media. Typically, citizens only have a vague and anecdotal understanding of the political orientation of newspapers and how this relates to voices on other platforms, such as social media. Undetected media bias can substantially impair democratic processes when uninformed citizens rely on such media to cast votes or make other politically relevant decisions. My analysis contributes to addressing such issues by giving a transparent, quantifiable, and analytically sound overview of the political orientation of newspapers and other media.
Secondly, my work promotes more balanced reporting in newspapers and other media. On the one hand, it might prompt newspapers to internally investigate and adjust their coverage. On the other hand, search engines and news aggregators might use my findings to implement features, such as tagging content that significantly deviates from the norm in terms of sentiment towards certain

political agents.

Moreover, the *NZZ* has been able to leverage insights from my sentiment and topic modeling analyses to publish an article on the reporting of Swiss newspapers throughout Covid-19.

4.4 Webpage

An interactive presentation of the results for the newspaper analysis, Twitter analysis, and the comparative analysis can be found on facesofthepandemic.

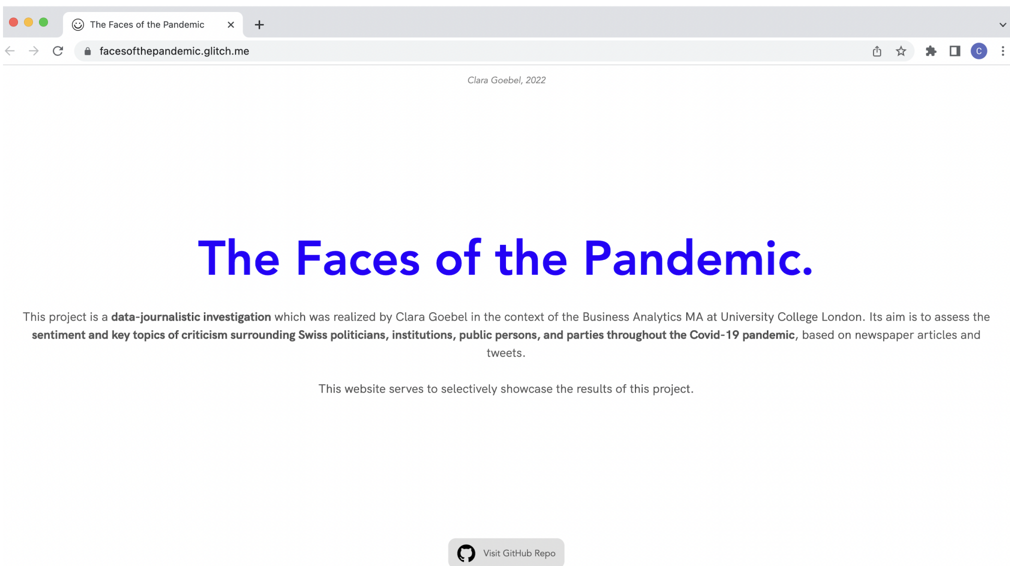

Figure 25: **Landing page:** The reader is presented with a brief overview of the project.

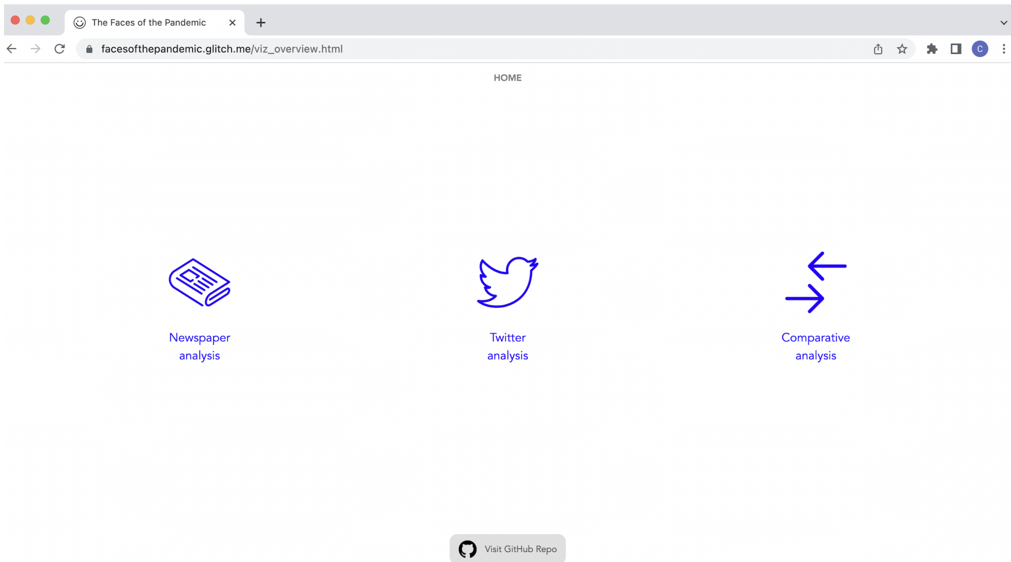

Figure 26: **Toggle menu:** The reader can choose which part of the analysis to read more about.

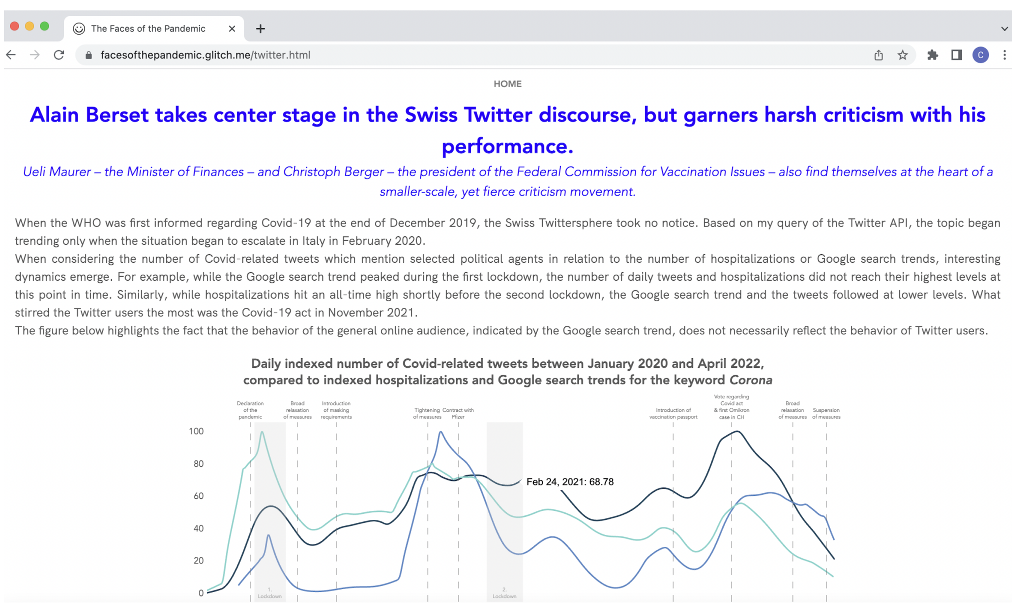

Figure 27: **Exemplary blog entry for the Twitter analysis:** The reader can scroll through in-depth interpretations and interactive charts of the results.

5 Conclusion

In this project, I have shed light on the impartiality of the Covid-19 coverage in Switzerland using natural language processing methods. In particular, it was analyzed i) to what extent the mentions of various political agents in Covid-related newspaper articles and tweets are sentiment-laden, ii) what are salient topics of criticism, and iii) how this evolved throughout January 2020 to April 2022. To this end, I drew on a wide range of programming languages (including Python, HTML, CSS, and Javascript), tools (including Snorkel), and data science methods (including machine learning, deep learning, timeseries analysis, topic modeling, and data visualization). With an accuracy of over 80%, I generated reliable findings on the status quo of political sentiment in the Swiss media and made important sociopolitical and company-specific contributions.

Nonetheless, my project still faces several limitations in its current state. Firstly, when identifying the agents in the newspaper articles, coreference resolution could be implemented to enable a more precise detection of agents. Currently, only sentences are picked up, where an agent is explicitly mentioned by name. Coreference resolution would allow me to identify sentences where agents are mentioned implicitly through referring expressions, such as pronouns. Unfortunately, coreference resolution methods for the German language are not fully operational yet, although prototypes exist, for example by Tuggener [76].
Furthermore, given the strict confidentiality of my dataset from the *SMD*, I was not able to use cloud-based computing resources which offer a GPU, such as Amazon Web Services. This implied computational constraints, for example when using embeddings or training transformer models, such as BERT. Provided broader computational resources, the embeddings and models could have been finetuned more machine learning and deep learning methods extensively and, potentially, achieved better results.

In terms of potential future work branching off of this project, several approaches show promise. For example, the scope of the analysis could be broadened by not only considering a small set of agents in the context of Covid-19, but analyzing newspapers and other media over a longer-term time horizon and also considering companies, NGOs, religious movements, intergovernmental alliances, and other entities. This would allow to granularly plot newspapers onto a compass of sociopolitical orientation, for example along the classic axes left vs. right and authoritarian vs. libertarian.
Another approach would be to improve the presentation of the results to the public by providing a real-time barometer of the political orientation of newspapers, with the option to filter by topic (e.g. Covid-19) and agent (e.g. Alain Berset).

6 Appendix

6.1 Obtaining the newspaper dataset

The dataset used for the newspaper analysis stems from the Swissdox database, which is made available by the *Linguistic Research Infrastructure (LiRI) of the University of Zurich* and the *Swiss Media Database (SMD)* [46]. This database contains approximately 29 million media articles from a wide range of Swiss media sources.
The dataset was obtained via the following parameters:

Parameter	**Input**
Language	German
Timeframe	01.01.2022 to 12.04.2022
Sources	All
Keywords	The keywords were searched for in the header and subheader of each article. • **covid** • **corona** • **pandemie** • **sars-cov** • **lockdown** • **maske** • **hospitali** • **impf** • **omikron** • **quarant** • **booster** • **contact-tracing** • **PCR** • **antigen** • **schutzkonzept** • **superspread** • **inzidenz**

Table 9: **Query parameters to obtain newspaper dataset**

The resulting dataset contains approximately 87'000 media articles from 48 sources. For each article, the following variables are recorded:

Variable	Description
`id`	Unique identifier of article
`pubtime`	Date and time of article release
`medium_code`	Code for media source
`medium_name`	Name of media source
`rubric`	Rubric of article
`regional`	Region of media source
`doctype`	Code for type of media source (e.g. PRD)
`doctype_description`	Name for type of media source (e.g. regional newspaper)
`language`	Language of article
`char_count`	Character count of article
`dateline`	Dateline of article
`head`	Title of article
`subhead`	Subtitle of article
`content_id`	Unique identifier of article content
`content`	Content of article

Table 10: **Variables contained in original newspaper dataset from Swissdox by LiRI**

6.2 Obtaining the Twitter dataset

The dataset used for the Twitter analysis was gleaned through the Twitter API via an academic research access. It was obtained via the following parameters:

Parameter	Input
Language	German
Timeframe	01.01.2022 to 12.04.2022
Retweets	Excluded

Query	The query was built from all possible combinations between a selection of agents and Covid-related keywords. The keywords for the agents consist of their name, where applicable their Twitter handle, and exclusions of any German or Austrian counterparts. • *"Ueli Maurer"* • *(Parmelin OR @ParmelinG)* • *(Berset OR @alain_berset)* • *(Bund OR Bundesrat) -Merkel -Scholz -Kurz -Nehammer -Bundeskanzler -Kanzler* • *(BAG OR "Bundesamt für Gesundheit" OR @BAG_OFSP_UFSP)* • *(Taskforce OR @SwissScience_TF) -Scheuer -Spahn -Anschober -Mückstein* • *"Daniel Koch"* • *Mathys* • *"Christoph Berger"* • *"Lukas Engelberger"* • *("Tanja Stadler" OR @TanjaStadler_CH)* • *Covid* • *Corona* • *Pandemie* • *Lockdown* • *Maske* • *Masken* • *Maskenpflicht* • *Impfung* • *Impfen* • *Impfpflicht* • *Omikron* • *Quarantäne* • *Booster* • *"Contact-Tracing"*

Table 11: **Query parameters to obtain Twitter dataset**

One challenge occurred when searching for tweets mentioning political bodies or institutions that also exist in Germany or Austria, such as *Taskforce*, as the aim was to return tweets referring to Switzerland only. Hence, key German and Austrian politicians, such as Germany's health minister *Karl Lauterbach*, were excluded in the API query. Moreover, the dataset was later filtered for keywords, such as mentions of cities or currencies, to improve the accuracy.

The resulting dataset contains approximately 257'000 tweets. For each tweet, the following variables are recorded:

Variable	**Description**
`id`	Unique identifier of tweet
`conversation_id`	Unique identifier of conversation, of which the tweet may be part of
`in_reply_to_user_id`	Unique identifier of user, to which the tweet may respond
`text`	Tweet content
`keyword`	Covid-related keyword referenced in tweet
`entity`	Agent referenced in tweet
`retweet_count`	Number of retweets
`reply_count`	Number of replies
`like_count`	Number of likes
`quote_count`	Number of tweet quotes
`created_at`	Date and time of tweet publication
`user_id`	Unique identifier of tweet author
`user_name`	User's name as displayed on profile
`user_username`	User's username as selected during registration
`user_verified`	Indicator of whether user is verified
`user_location`	User's location as displayed on profile
`user_followers_count`	Number of accounts following the user
`user_following_count`	Number of accounts user is following
`user_tweet_count`	Number of tweets authored by the user
`user_listed_count`	Number of lists a user has been added to
`user_geo`	User geography as specified during registration
`user_country_code`	User country as specified during registration

Table 12: **Variables contained in original Twitter dataset**

6.3 Agents tested for in newspaper articles

Keyword	**Designed Agent**	**Associated Group**	**Affiliated Party**	**Agent Type**
`Maurer`	Ueli Maurer	Bundesrat	SVP	Politician
`Parmelin`	Guy Parmelin	Bundesrat	SVP	Politician
`Cassis`	Ignazio Cassis	Bundesrat	FDP	Politician
`Sutter`	Karin Keller Sutter	Bundesrat	FDP	Politician
`Sommaruga`	Simonetta Sommaruga	Bundesrat	SP	Politician
`Berset`	Alain Berset	Bundesrat	SP	Politician
`Amherf`	Alain Berset	Bundesrat	SP	Politician
`bundesr\w*`	Bundesrat	Bundesrat	n/a	Political body

`\\b[dD]\w*`
`Bund[es]{0,2}`
`\\b`
	Bundesrat	Bundesrat	n/a	Political body

`nationalr\w*`	Nationalrat	Nationalrat	n/a	Political body
`staender\w*`	Ständerat	Ständerat	n/a	Political body
`Stadler` `(?!\ Rail)`	Tanja Stadler	Taskforce	n/a	Official
`(?<!Simon\ )` `Tanner`	Marcel Tanner	Taskforce	n/a	Official
`\w*taskforce`	Taskforce	Taskforce	n/a	Institution
`Berger`	Christoph Berger	EKIF	n/a	Official
`EKIF`	EKIF	EKIF	n/a	Institution
`Kuster`	Stefan Kuster	BAG	n/a	Official

Masserey	Virginie Masserey	BAG	n/a	Official
Mathys	Patrick Mathys	BAG	n/a	Official
(?<!Robert\ ) Koch	Daniel Koch	BAG	n/a	Official
BAG	BAG	BAG	n/a	Institution
bundesamt[es]{0,2} fuer gesundheit	BAG	BAG	n/a	Institution
swissmedic	Swissmedic	Swissmedic	n/a	Institution
Engelberger	Lukas Engelberger	GDK	Die Mitte	Official
GDK	GDK	GDK	n/a	Institution
gesundheitsdirektorenkonferenz	GDK	GDK	n/a	Institution
SVP\\b	SVP	SVP	SVP	Party
SP\\b	SP	SP	SP	Party
FDP\\b	FDP	FDP	FDP	Party
CVP\\b	Die Mitte	Die Mitte	Die Mitte	Party
\\b[dD]\w* Mitte\\b	Die Mitte	Die Mitte	Die Mitte	Party
\\b[dD]\w* Gruene\w*	Die Grüne	Die Grüne	Die Grüne	Party
Gruenliberale\w*	Grünliberale	Grünliberale	Grünliberale	Party
Juso	Juso	Juso	Juso	Party
\w*befuerwort\w*	Befürworter	Approving	n/a	Political camp
ja lager	Ja Lager	Approving	n/a	Political camp
\w*gegner\w*	Gegner	Disapproving	n/a	Political camp

\w*leugner\w*	Leugner	Disapproving	n/a	Political camp
\w*skeptiker\w*	Skeptiker	Disapproving	n/a	Political camp
\w*kritiker\w*	Kritiker	Disapproving	n/a	Political camp
opposition	Opposition	Disapproving	n/a	Political camp
nein lager	Nein Lager	Disapproving	n/a	Political camp
\w*demon-strant\w*	Demonstranten	Disapproving	n/a	Political camp
freunde\w* der verfassung	Freunde der Verfassung	Disapproving	n/a	Political camp
mass\\b voll\\b	Mass Voll	Disapproving	n/a	Political camp

Table 13: **Agents tested for in newspaper articles**

6.4 Finalized newspaper dataset

Variable	Description
id	Unique identifier of article
pubtime	Date and time of article release
pubday	Date of article release
pubmonth	Month of article release
medium_name	Name of media source, where versions (e.g. online vs. offline editions, weekday vs. Sunday editions, etc.) are aggregated
doctype_description	Name for type of media source (e.g. regional newspaper)
channel_description	Channel where article was published
char_count	Character count of full content of article

`original_content`	Full article content, including title, subtitle, and article content with mention of agent or multiple agents
`original_sentence`	Corresponding sentence from title, subtitle, or article content with mention of agent or multiple agents
`original_passage`	Corresponding passage, including one sentence before and after sentence with mention of agent or multiple agents
`sentence_ABSA`	Lowercase sentence, where structural changes for ABSA (e.g. removal of negatively connoted agents) have been performed
`passage_ABSA`	Lowercase passage, where structural changes for ABSA (e.g. removal of negatively connoted agents) have been performed
`sentence_ABSA_rel_keywords`	Parsed version of `sentence_ABSA`, where only tokens that are grammatically closely related to the agent are selected
`sentence_ABSA_subclause`	Parsed version of `sentence_ABSA`, where only subclauses in which the agent is mentioned are selected
`clause_ABSA`	Combination of `sentence_ABSA` and `sentence_ABSA_subclause`, where `sentence_ABSA_subclause` is selected in case multiple agents are contained in `original_sentence`
`entity_name`	Designed agent contained in clause
`entity_keyword`	Regex match for agent contained in clause
`num_entities`	Number of agents contained in `original_sentence`

Table 14: **Variables contained in final newspaper dataset**

6.5 Finalized Twitter dataset

Variable	**Description**
`id`	Unique identifier of tweet

`conversation_id`	Unique identifier of conversation, of which the tweet may be part of
`in_reply_to_user_id`	Unique identifier of user, to which the tweet may respond
`original_text`	Unmodified tweet content
`text`	Cleaned tweet content
`emojis`	Emojis contained in tweet
`text_cleaned_ emojis_replaced`	Cleaned tweet content with emojis replaced with words indicative of the sentiment expressed by the emoji
`char_count`	Character count of tweet text
`keyword`	Covid-related keyword referenced in tweet
`entity`	Agent referenced in tweet
`reply`	Indicator of whether tweet is a reply
`conversation`	Indicator of whether tweet forms part of a conversation
`retweet_count`	Number of retweets
`reply_count`	Number of replies
`like_count`	Number of likes
`quote_count`	Number of tweet quotes
`pubtime`	Date and time of tweet publication
`pubday`	Date of tweet publication
`pubmonth`	Month of tweet publication
`user_id`	Unique identifier of tweet author
`user_name`	User's name as displayed on profile
`user_username`	User's username as selected during registration
`user_verified`	Indicator of whether user is verified
`user_location`	User's location as displayed on profile
`user_followers_count`	Number of accounts following the user
`user_following_count`	Number of accounts user is following
`user_tweet_count`	Number of tweets authored by the user

user_listed_count	Number of lists a user has been added to

Table 15: **Variables contained in final Twitter dataset**

References

[1] Amazon. Covid-19 News Sentiment Analyzer [online]. 2020. URL: https://aws.amazon.com/marketplace/pp/prodview-u7mhbuegx4yvw.

[2] M. Ansaria, M. Aziza, M. Siddiquib, H. Mehraa, and K. Singha. Analysis of Political Sentiment Orientations on Twitter. *International Conference on Computational Intelligence and Data Science (ICCIDS)*, 2019.

[3] S. Ansolabehere, R. Lessem, and J. Snyder. The Orientation of Newspaper Endorsements in U.S. Elections. *Quarterly Journal of Political Science*, 1(4):393–404, 2006.

[4] F. Aslam, T. Awan, J. Syed, A. Kashif, and M. Parveen. Sentiments and Emotions Evoked by News Headlines of Coronavirus Disease (Covid-19) Outbreak. *Humanities and Social Sciences Communications*, 7:23, 2020.

[5] A. Balahur and R. Steinberger. Rethinking Sentiment Analysis in the News: From Theory to Practice and Back. *European Commission - Joint Research Centre*, 2009.

[6] P. Barbera. Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data. *Political Analysis*, 2014.

[7] A. Bermingham and A. Smeaton. On Using Twitter to Monitor Political Sentiment and Predict Election Results. *Proceedings of the Workshop on Sentiment Analysis where AI Meets Psychology (SAAIP)*, pages 2–10, 2011.

[8] D. Blei, A. Ng, and M. Jordan. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.

[9] S. Boon-Itt and Y. Skunkan. Public Perception of the Covid-19 Pandemic on Twitter: Sentiment Analysis and Topic Modeling Study. *JMIR Public Health and Surveillance*, 6(4):1–17, 2020.

[10] F. Bornebusch, G. Cancino, A. Fanseu, M. Jalali, J. Mohsen, M. Nitze, C. Plump, R. Yotedje, H. Tchambo, and H. Ziegler. *Informatik*. Gesellschaft für Informatik, 2014.

[11] A. Boutet, H. Kim, and E. Yoneki. What's in Your Tweets? I Know Who You Supported in the U.K. 2010 General Election. *Proceedings ICWSM*, 2012.

[12] L. Breiman. Random Forests. *Machine Learning*, 45:5–32, 2001.

[13] T. Brenn. SRF-Chefredaktor nimmt Stellung zu Fakenews-Vorwürfen. *SRF*, 2020. URL: https://www.srf.ch/news/schweiz/berichterstattung-zu-corona-srf-chefredaktor-nimmt-stellung-zu-fakenews-vorwuerfen.

[14] K. Chakraborty, S. Bhatia, S. Bhattacharyya, J. Platos, R. Bag, and Ab. Hassanien. Sentiment Analysis of Covid-19 Tweets by Deep Learning Clas-

sifiers: A Study to Show how Popularity is Affecting Accuracy in Social Media. *Applied Soft Computing Journal*, 97, 2020.

[15] T. Chen and C. Guestrin. XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794, 2016.

[16] R. Cohen and D. Ruths. Classifying Political Orientation on Twitter: It's Not Easy! *Proceedings ICWSM*, 2013.

[17] M. Conover, B. Goncalves, J. Ratkiewicz, A. Flammini, and F. Menczers. Predicting the Political Alignment of Twitter Users. *IEEE International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing*, pages 192–199, 2011.

[18] M. Costola, M. Nofer, O. Hinz, and L. Pelizzon. Machine Learning Sentiment Analysis, Covid-19 News and Stock Market Reactions. *Leibniz Institute for Financial Research SAFE*, 288, 2020.

[19] J. Dehler-Holland, M. Okoh, and D. Keles. Assessing Technology Legitimacy with Topic Models and Sentiment Analysis - The Case of Wind Power in Germany. *Technological Forecasting Social Change*, 175, 2022.

[20] J. Devlin, M. Chang, K. Lee, and K. Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of NAACL-HLT*, pages 4171–4186, 2019.

[21] M. Eisenegger, F. Oehmer, L. Udris, and D. Vogler. Die Qualität der Medienberichterstattung zur Corona-Pandemie. *Forschungszentrum Öffentlichkeit und Gesellschaft / Universität Zürich*, 2020.

[22] M. Eisenegger, F. Oehmer, L. Udris, and D. Vogler. Lessons Learned? Die Qualität der Medienberichterstattung in der ersten und zweiten Welle der Corona-Pandemie. *Forschungszentrum Öffentlichkeit und Gesellschaft / Universität Zürich*, 2021.

[23] F. Falck, J. Marstaller, N. Stoehr, S. Maucher, J. Ren, A. Thalhammer, A. Rettinger, and R. Studer. Sentiment Political Compass: A Data-driven Analysis of Online Newspapers regarding Political Orientation. *The Internet, Policy Politics Conference*, 2018.

[24] J. Fehle, T. Schmidt, and C. Wolff. Lexicon-based Sentiment Analysis in German: Systematic Evaluation of Resources and Preprocessing Techniques. *Proceedings of the 17th Conference on Natural Language Processing*, pages 86–103, 2021.

[25] Bundesamt für Gesundheit (BAG). Laboratory-confirmed Hospitalizations [online]. 2022. URL: https://www.covid19.admin.ch/en/epidemiologic/hosp.

[26] F. Gaisbauer, A. Pournaki, S. Banisch, and E. Olbrich. Ideological Differences in Engagement in Public Debate on Twitter. *PLoS ONE*, 16(3), 2021.

[27] K. Garcia and L. Berton. Topic Detection and Sentiment Analysis in Twitter Ccontent Related to Covid-19 from Brazil and the U.S. *Applied Soft Computing Journal*, 101, 2021.

[28] M. Gentzkow and J. Shapiro. What drives Media Slant? Evidence from U.S. Daily Newspapers. *Econometrica*, 78(1):35–71, 2010.

[29] P. Giehl. Automated Sentiment Analysis of the Swiss Media Landscape. Master's thesis, 2018.

[30] F. Gilardi, T. Gessler, M. Kubli, and S. Müller. Social Media and Policy Responses to the Covid-19 Pandemic in Switzerland. *Swiss Political Science Review*, 27:243–256, 2021.

[31] Google. Google Search Trend [online]. 2022. URL: https://trends.google.com/trends/?geo=CH.

[32] O. Guhr, A. Schumann, F. Bahrmann, and H. Böhme. Training a Broad-Coverage German Sentiment Classification Model for Dialog Systems [online]. 2020. URL: https://huggingface.co/oliverguhr/german-sentiment-bert.

[33] D. Ho and K. Quinn. Measuring Explicit Political Positions of Media. *Quarterly Journal of Political Science*, 3(4):353–377, 2008.

[34] M. Hoang, O. Bihorac, and J. Rouces. Aspect-Based Sentiment Analysis Using BERT. *Proceedings of the 22nd Nordic Conference on Computational Linguistics (NoDaLiDa)*, pages 187–196, 2019.

[35] A. Hossain, M. Karimuzzaman, M. Hossain, and A. Rahman. Text Mining and Sentiment Analysis of Newspaper Headlines. *Information*, 12:414–429, 2021.

[36] J. Huang, Y. Meng, Gua F, H. Ji, and J. Han. Weakly-Supervised Aspect-Based Sentiment Analysis via Joint Aspect-Sentiment Topic Embedding. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, pages 6989–6999, 2020.

[37] A. Imran, S. Daudpota, Z. Kastrati, and R. Batra. Cross-Cultural Polarity and Emotion Detection Using Sentiment Analysis and Deep Learning on Covid-19 Related Tweets. *IEEE Transactions on Computational Social Systems*, 8:181074, 2020.

[38] Snorkel AI Inc. Snorkel [online]. 2015. URL: https://snorkel.ai/.

[39] M. Kaya, G. Fidan, and I. Toroslu. Sentiment analysis of turkish political news. *International Conferences on Web Intelligence and Intelligent Agent Technology*, 2012.

[40] A. Khatua and E. Cambria. Predicting Political Sentiments of Voters from Twitter in Multi-Party Contexts. *Applied Soft Computing Journal*, 97, 2020.

[41] F. Kinzelmann, P. Wallimann, and R. Röthlin. So gut meisterten wir Corona. *Blick*, 2022. URL: https://www.blick.ch/ausland/bilanz-der-krisenbewaeltigung-die-schweiz-liegt-im-weltweiten-vergleich-auf-platz-2-so-html.

[42] M. Klašnja, P. Barberá, N. Beauchamp, J. Nagler, and J. Tucker. *Measuring Public Opinion with Social Media Data*, pages 555–582. The Oxford Handbook of Polling and Survey Methods, 2015.

[43] The Digital Democracy Lab. Wie die Twitter-Debatte zum Coronavirus in der Schweiz Überhand Nimmt. 2020. URL: https://digdemlab.io/eye/2020/04/01/coronavirus.html?fbclid=IwAR3XsnixTY-hbHFGnjlyzsHmA3TawaE77DzFrGUwur4L9QqMNK8ihSU2fus#.

[44] D. Lee and H. Seung. Algorithms for Non-Negative Matrix Factorization. *Advances in Neural Information Processing Systems*, 13, 2000.

[45] J. Li. From Tweets and Newspapers to Polls: A Sentiment Study on the 2017 U.K. General Election. Master's thesis, 2018.

[46] LiRI. Swissdox@LiRI [online]. 2022. URL: https://www.liri.uzh.ch/en/services/swissdox.html.

[47] S. Loria and M. Killer. Textblob-de [online]. 2014. URL: https://textblob-de.readthedocs.io/en/latest/.

[48] S. Lüdke, J. grau, and M. Drawitsch. German Sentiment BERT Finetuned on News Data [online]. 2020. URL: https://huggingface.co/mdraw/german-news-sentiment-bert.

[49] S. Lüdke, J. grau, and M. Drawitsch. News Sentiment Development on the Example of "Migration". 2020. URL: https://github.com/text-analytics-20/news-sentiment-development.

[50] M. MacKay, T. Colangeli, S. Gosselin, S. Neumann, and A. Papadopoulos. Engagement Analysis of Canadian Public Health and News Media Facebook Posts and Sentiment Analysis of Corresponding Comments during Covid-19. *Psych*, 4:60–70, 2022.

[51] L. Mascarell, T. Ruzsics, C. Schneebeli, P. Schlattner, L. Campanella, S. Klingler, and C. Kadar. Stance Detection in German News Articles. *Proceedings of the Fourth Workshop on Fact Extraction and VERification (FEVER) at EMNLP*, pages 66–77, 2021.

[52] J. Mellon and C. Prosser. Twitter and Facebook are not Representative of the General Population: Political Attitudes and Demographics of British Social Media Users. *Research Politics*, 4(3), 2017.

[53] C. Müller, B. Wüest, and T. Willi. Twitter Data for Political Analysis in Switzerland. *Swiss Political Science Association*, 2016.

[54] S. Na, L. Xumin, and G. Yong. Research on K-means Clustering Algorithm: An Improved K-means Clustering Algorithm. *Third International Symposium on Intelligent Information Technology and Security Informatics*, 2010.

[55] D. Oelke, B. Geisselmann, and D. Keim. Visual Analysis of Explicit Opinion and News Bias in German Soccer Articles. *International Workshop on Visual Analytics*, 2012.

[56] Kralj Novak P., Smailović J., Sluban B., and Mozetič I. Sentiment of Emojis [online]. 2015. URL: https://pypi.org/project/emosent-py/.

[57] C. Peng, K. Lee, and G. Ingersoll. An Introduction to Logistic Regression Analysis and Reporting. *The Journal of Educational Research*, 96(1):3–14, 2010.

[58] H. Peng, L. Xu, L. Bing, F. Huang, W. Lu, and L. Si. Knowing What, How and Why: A Near Complete Solution for Aspect-Based Sentiment Analysis. *The Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-20)*, 2020.

[59] A. Ratner, S. Bach, H. Ehrenberg, J. Fries, S. Wu, and C. Ré. Snorkel: Rapid Training Data Creation with Weak Supervision. *Proceedings of the VLDB Endowment*, 11(3), 2017.

[60] A. Ratner, B. Hancock, J. Dunnmon, F. Sala, S. Pandey, and C. Ré. Training Complex Models with Multi-Task Weak Supervision. *The Thirty-Third AAAI Conference on Artificial Intelligence*, 2019.

[61] R. Remus, K. Ahmad, and G. Heyer. Sentiment in German-Language News and Blogs, and the DAX. *Leipziger Beitrage zur Informatik*, pages 149–158, 2009.

[62] R. Remus, U. Quasthoff, and G. Heyer. SentiWS - A Publicly Available German-Language Resource for Sentiment Analysis [online]. 2010. URL: https://spacy.io/universe/project/spacy-sentiws.

[63] A. Rostetter. ≪Mister Corona≫ will die Bühne nicht verlassen. *NZZ*, 2020. URL: https://www.nzz.ch/schweiz/der-ehemalige-mister-corona-spielt-eine-zunehmend-fragwuerdige-rolle-ld.1569992?reduced=true.

[64] F. Rustam, M. Khalid, W. Aslam, V. Rupapara, A. Mehmood, and G. Choi. A Performance Comparison of Supervised Machine Learning Models for Covid-19 Tweets Sentiment Analysis. *PLoS ONE*, 16(2):1–23, 2021.

[65] J. Samuel, G. Ali, M. Rahman, E. Esawi, and Y. Samuel. Covid-19 Public Sentiment Insights and Machine Learning for Tweets Classification. *Information*, 11:314–336, 2020.

[66] A. Sherstinsky. Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network. *Physica D: Nonlinear Phenomena*, 404, 2019.

[67] C. Shofiya and S. Abidi. Sentiment Analysis on Covid-19-Related Social Distancing in Canada Using Twitter Data. *International Journal of Environmental Research and Public Health*, 18:5993, 2021.

[68] L. Shu, H. Xu, B. Liu, and J. Chen. Zero-Shot Aspect-Based Sentiment Analysis. *Forthcoming*, 2022.

[69] SRG. SRG Wahlbarometer [online]. 2021. URL: https://www.srf.ch/news/schweiz/srg-wahlbarometer-welche-bundes/raete-gelten-als-einflussreich-und-sympathisch.

[70] SRG. 10. SRG Corona-Umfrage [online]. 2022. URL: https://sotomo.ch/site/wp-content/uploads/2022/02/10.-SRG-Corona-Monitor-Sotomo.pdf.

[71] S. Stöckl. Wirbel um Aussage von Ringier-Chef: ≪Regierungstreu berichten≫. *SRF*, 2022. URL: https://www.srf.ch/audio/heutemorgen/wirbel-um-aussage-von-ringier-chef-regierungstreu-berichten?id=12117476.

[72] Swissinfo. Nationalratskommission präzisiert Maulkorb für Taskforce. *Swissinfo*, 2021. URL: https://www.swissinfo.ch/ger/nationalratskommission-praezisiert-maulkorb-fuer-taskforce/46423776.

[73] Y. Tian, G. Chen, and Y. Song. Enhancing Aspect-level Sentiment Analysis with Word Dependencies. *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics*, pages 3726–3739, 2021.

[74] S. Tong and D. Koller. Support Vector Machine Active Learning with Applications to Text Classification. *Journal of Machine Learning Research*, (45-66), 2001.

[75] NLP Town. BERT Base Multilingual Uncased Sentiment [online]. 2020. URL: https://huggingface.co/nlptown/bert-base-multilingual-uncased-sentiment.

[76] D. Tuggener. *Incremental Coreference Resolution for German*. PhD thesis, 2016.

[77] L. van der Maaten and G. Hinton. Visualizing Data Using t-SNE. *Journal of Machine Learning Research*, 9:2579–2605, 2008.

[78] F. Wong, C. Tan, S. Sen, and M. Chiang. Quantifying Political Leaning from Tweets, Retweets, and Retweeters. *IEEE Transactions on Knowledge and Data Engineering*, 2016.

[79] J. Zhang, Y. Yu, Y. Li, Y. Wang, Y. Yang, M. Yang, and A. Ratner. WRENCH: A Comprehensive Benchmark for Weak Supervision. *35th Conference on Neural Information Processing Systems (NeurIPS)*, 2021.

[80] Z. Zhou. A Brief Introduction to Weakly Supervised Learning. *National Science Review*, 5:44–53, 2018.