A EVALUATION OF DATAVOIDANT

A.1 Evaluation of the Machine Learning Algorithms in Datavoidant.

Our system uses state-of-the-art machine learning models that learn how to categorize social media data to help end-users identify data voids. We study whether the automated approaches that we utilize match human intuition, specifically how humans themselves would categorize the social media data.

A.1.1 Dataset for Studying Machine Learning Models. To support the evaluation of our machine learning algorithms, we created a dataset. We use the dataset to help us have a way for comparing the categorization conducted by our machine learning algorithms to how humans would categorize the same data. For this purpose, we asked 10 journalists to first provide a list of Facebook groups and pages from underrepresented communities for which they would like to potentially study and address data voids. The 10 journalists provided a list of 1,150 Facebook groups and pages. Next, we collected a month's worth of data from the groups and pages, collecting a total of 271,717 Facebook posts.

A.1.2 Evaluation of the Topic Categorization Machine Algorithms. We believe that our system will be more intuitive and better for independent journalists to use, the more the system's automation matches human decision-making. We therefore, compared the topic categorization of our system to the topic categorization done by humans. For this purpose, we first asked the 10 journalists who helped us to create our initial dataset (See A.1.1) to provide the number of topics that they wanted to consider for studying data voids (recall this is one of the minimal inputs that our system needs for the categorization of content). The journalists agreed on having 11 topics to study the data voids. They defined the number of topics based on the number of issues that the Pew Research Center reported as mattering the most to underrepresented communities in the 2020 US presidential election [86]. The journalists considered they wanted to cover data voids within these 11 key issues. Next, we asked three of the journalists to categorize a subset of posts in our dataset independently into the 11 topics. For this purpose, we used stratified sampling to collect 5% of the Facebook posts from our dataset ensuring the posts covered all 11 topics. Next, we asked two of the coders to manually categorize each of the 13,585 posts using one of the 11 topics. We asked the workers to pick the "most relevant" topic for each post. The two workers agreed on 81.82% posts (Cohen's kappa: .80). We then asked the third coder to label the remaining posts upon which the first two coders had disagreed. We then used a "majority rule" approach to determine the topic for those posts. At the end of this step, we had all the posts of our dataset categorized into one of the 11 topics. We considered this human categorized dataset to be our "Gold Standard". Next, we separated this Gold Standard dataset into 80% and 20% for the training and validation of our system's data categorization. We evaluate our approach on the validation set. We had the following *results*: 208,869 / 271,717 = 76.87% accuracy (with 11 topics). These results suggest that the machine learning algorithm of our system can successfully categorize posts into topics that are similar to how humans would do the categorization.

A.1.3 Evaluation of the Political Leaning Categorization. Our system has a module that automatically categorizes content into its political leaning (primarily "conservative" or "liberal"). We are interested in studying how accurate this automated categorization is, especially in comparison to how humans would categorize the political leaning of the same content. For this purpose, we used stratified sampling to collect 5% of the Facebook posts from our dataset, ensuring the posts covered all 11 topics and also had a balance of Facebook groups and pages from citizens, political actors, and news media outlets. We then asked the three journalists who had done the topic categorization to help us again to conduct the categorization of the political leaning of the posts. We asked two of

the coders to categorize each of the 13,585 posts into whether they were "liberal" or "conservative". We asked them to take into account if the post mentioned a political actor, the website political leaning score from Robertson et al. [125] dataset, and the tone (sentiment) of the posts pick the "most relevant" political leaning for each post. If the post mentioned neither websites nor political actors we asked them to classify it as neutral. The two coders agreed on 80% of posts (Cohen's kappa: 0.70). We then asked the third coder to label the posts upon which the first two coders had disagreed. We again used a "majority rule" approach to determine the political leaning of those posts. After this step, we had a dataset with "gold-standard" labels of the political leanings. Armed with our dataset, we tested how much our algorithm could accurately classify posts into their political leanings according to the gold standard dataset. Our algorithm achieved a precision of 74.42%; recall of 94.12%; and accuracy of 81.43%. Details are in Table 2. This result suggests that our political leaning identification module can successfully categorize liberal and conservative posts. This helps the system to identify data voids that relate to political leanings.

	Real: liberal	Real: conservative	Precision	Recall	Accuracy	F1-score
Pred: liberal	32	11	74.42%	94.12%	81.43%	83.12%
Pred: conservative	2	25				

Table 2. Results of the classification of political leaning of posts

A.1.4 Evaluation of the Bot Detection Machine Learning Algorithm. We trained our machine learning algorithms that detect bots on the comprehensive bot detection benchmark of TwiBot-20 [49]. The benchmark provides a dataset that has manually categorized social media accounts into "bots" and "humans" (i.e., they provide a gold standard). We evaluate the machine learning algorithm that we use for detecting bots on the test set of [49]. Our algorithm achieved a precision of 76.09%; a recall of 86.19% and accuracy of 80.47%, which is comparable to other state-of-the-art bot detection algorithms. See details on Table 3. Given these results, we argue that our bot detection module enables our system to identify and present the online political narratives that automated accounts could push.

	Gold: bot	Gold: human	Precision	Recall	Accuracy	F1-Score
Pred: bot	487	153	80.47%	76.09%	86.19%	80.83%
Pred: human	78	465				

Table 3. Results of our Bot Detection Machine Learning Algorithm.

A.2 Smart Categorization Module

Given that the data collected by the *Data Collection Module* can be massive and difficult for humans to interpret, this module focuses on structuring and categorizing the data to facilitate collective sensemaking. For this purpose, Datavoidant uses state-of-the-art machine learning models to categorize social media content and then synthesize results (*"Step: Schematize"* in the sensemaking loop). First, Datavoidant uses basic NLP techniques to categorize the Facebook groups and pages into either "content from political actors," "content from citizen initiatives," or "content from news sites." In particular, the system uses public datasets that list different news sites [3], especially datasets of news sites targeting underrepresented populations [9], to analyze whether the name of a given Facebook group or page matches any of the news sites in the datasets. If it finds a match, the system labels that Facebook group or page as "content from news sites." For example, if a journalist inputs the Facebook page of *"The New York Times"* [1], the system will label that page

as being "content from news sites" because it found a match in the dataset. Similarly, to identify whether a Facebook page or group is "content from political actors," Datavoidant takes the name and description of the page, and analyzes whether it directly mentions the term *"political"* (or related synonyms). Datavoidant also analyzes whether the page mentions a political actor or political party in its name. For this purpose, Datavoidant crawls Wikipedia to obtain lists of political actors and political parties to consider[7, 8]. All other Facebook groups and pages are labeled as "content from citizen initiatives". Notice that we allow journalists to correct the system's categorization of Facebook groups and pages and re-categorize the content as they consider more appropriate. The system also allows journalists to input the sources of data they want Datavoidant to use for this first categorization (e.g., Wikipedia articles about political actors, lists of newspapers etc). After this step, we have all the Facebook groups and pages categorized into three main types: news media, political spaces, and citizen groups. This type of categorization was important given that journalists expressed an interest in being able to bridge the data gap between these different online spaces. However, it was also important for journalists to be able to conduct a multi-level analysis where they could understand what topics were covered less than others across these different online spaces, what political actors were pushing certain content, as well as identify whether automated methods were pushing certain topics (to understand manipulations around the data voids). For this purpose, Datavoidant integrates state-of-the-art machine learning models to categorize the content on multiple levels and facilitate these types of data analysis.

TOPIC LEVEL CATEGORIZATION. In the design of Datavoidant, we considered that journalists would likely not have the time or ability to interpret complex abstract topics without labels, like the ones that the topic modeling algorithm of LDA throws out [22]. We assume that most journalists will likely not know how to provide labeled data to train machine learning algorithms that can discern one topic from another. Therefore, we opted for automated methods that could remove the unnecessary burden and complexity to journalists, while still allowing them to automatically categorize their data at scale. Datavoidant simply asks journalists to provide the list of topics they are interested in exploring and a list of keywords associated with each topic. The system then uses these keywords and topics to automatically create a training and testing set to teach machine learning models how to classify posts into the different topics. In specific, for each topic, the system uses its associated keywords to randomly sample Facebook posts that mention the keywords. The posts are taken from the lists of Facebook groups and pages that the end-user provided initially. The system then labels each post with one topic, selecting the topic with the greatest number of keywords in the post. The system will aim to have the same number of posts for each topic, but allows the end user to know when this is is not the case. Through this, the end user can easily modify the topics and facilitate creating a more balanced dataset. Datavoidant then trains a pre-trained language model RoBERTa [94] and uses fully connected layers for topic classification. This model is trained on the collected, labeled dataset of Facebook posts and their topics with a 8:2 split for training and validation set.

POLITICAL LEANING CATEGORIZATION. In addition to topic-level data voids, Datavoidant also helps journalists to identify political-level data deficiencies, where some topics might be less discussed by accounts from certain political or ideological perspectives. For example, climate change content might be rarely covered by liberals, while critical race theory could be less covered by conservatives, creating partisan echo chambers and political-level data voids. For this purpose, Datavoidant identifies each post's political leaning to facilitate visualization and understanding of political-level data deficits. To conduct its automatic categorization of posts with respect to political leanings, Datavoidant resorts to external knowledge about the political leanings of websites [125] and political actors [47]. Datavoidant conducts the following approach to calculate the political leaning score of a given Facebook post:

- If the post comes from a Facebook page that represents a websites that is in the list (i.e., a website with a clear political leaning), the system:
  - averages the mentioned websites' political leaning score based on [125] and through this obtains the post's final "political leaning score" as b_w.
- If the post mentions any website on the list or mentions any political actors then the system:
  - calculates the sentiment score s [40, 149] for the post, with -1 as most negative and +1 as most positive.
  - averages the political leaning score of the actors and websites mentioned based on [47, 125] to obtain the "political leaning score" b_a. The final political leaning score of the post is then obtained by $b_a \times s$.
- If the post mentions neither websites nor political actors, the system takes 0 for its political leaning score and regard the post as neutral.

Notice that Datavoidant categorizes posts first based on the overall nature of the Facebook page from which the post is from. We consider that known conservative outlets will tend to always post conservative content and liberal outlets will tend to post liberal content. If the system cannot identify the nature of the Facebook page, it analyzes whether the post is discussing liberal or conservative actors in a positive or negative form, and uses this to calculate the political leaning score of the post. In all other cases, the system labels the post as neutral. In this way, Datavoidant calculates political leaning scores for social media posts, which helps to illustrate political-level data deficiencies across topics.

BOT CATEGORIZATION. Automated social media users, also known as bots, widely exist on online social networks and induce undesirable social effects. In the past decade, malicious actors have launched bot campaigns to interfere with elections [39, 52], spread misinformation [48] and propagate extreme ideology [20]. To these issues, Datavoidant includes a bot detection component that categorizes accounts into bots and none-bots. The aim is to help journalists identify biased information propagated by malicious actors. In Datavoidant, we focus on the textual content of posts to identify Facebook bots and malicious actors. Specifically, we follow the method in the state-of-the-art approach [50] to encode post content with pre-trained language models [94] and train a multi-layer perceptron for bot detection. We train our model with the comprehensive benchmark TwiBot-20 [49].

B PARTICIPANTS INTERVIEW STUDY

For the purpose of protecting the anonymity of our interviewees, we have anonymized the data from the journalists we recruited for our interview study. We followed guidelines used in prior work for disclosing information about journalists who take part in interview studies [70, 99].

Independent Journalist	Organization Type	Language
J1	Niche Newspaper	Monolingual
J2	Niche Newspaper	Bilingual
J3	Niche Newspaper	Monolingual
J4	Niche Newspaper	Monolingual
J5	Niche Newspaper	Monolingual
J6	Niche Newspaper	Monolingual
J7	U.S. Local Radio	Bilingual
J8	Niche Newspaper	Monolingual
J9	Niche Newspaper	Monolingual
J10	Non-Profit Newsroom & Civic Engagement Organization	Monolingual
J11	Niche Newspaper	Monolingual
J12	Digital First Outlet	Monolingual
J13	Digital First Outlet	Monolingual
J14	Non-Profit Newsroom & Civic Engagement Organization	Bilingual
J15	Digital First Outlet	Monolingual
J16	Digital First Outlet	Monolingual
J17	Digital First Outlet	Bilingual
J18	Digital First Outlet	Monolingual
J19	Digital First Outlet	Monolingual
J20	Digital First Outlet	Monolingual
J21	Digital First Outlet	Monolingual
J22	Digital First Outlet	Monolingual

Table 4. Overview of participants in our interview study