

A Appendix

A.1 Ethics and Reddit user anonymity

We acknowledge that even anonymous use of social media poses potential risk due to accidentally revealing personal information. For this reason, we would like to maintain a similar policy to `pushshift` which enables users to request that their posts / usernames be removed from the database. We disclose no information about users and ensure all data is used in compliance with reddit’s data policies.

A.2 Data pre-processing and filtering

As mentioned in section 3, certain pre-processing steps are taken to ensure we produce a dataset with high-quality textual interactions. In tables 1 and 2, we present the share of data each filter drops at the post- and comment-level respectively.

Table 1: Filtering to high-quality interactions - posts

	Subreddit				
	Brexit	Republican	Democrats	Climate	BLM
Posts with <k comments*	58%	71%	71%	75%	70%
Posts with <10 words	32%	44%	43%	30%	52%

where $k = 2, 5, 5, 5, 10$ for `r/BlackLivesMatter`, `r/climate`, `r/Republican`, `r/democrats`, and `r/Brexit`, respectively.

Table 2: Filtering to high-quality interactions - comments

	Subreddit				
	Brexit	Republican	Democrats	Climate	BLM
Comments with <10 words	21%	31%	32%	21%	34%
Comments with >100 words	12%	5%	5%	14%	6%
Comments with hyperlinks	6%	6%	4%	14%	6%
Comments with nest level >2	55%	42%	41%	45%	35%

A.3 Labeling instructions

We include the instructions given to annotators. Figures 1, 2, 3 contain the instructions for annotation. Annotators also have access to helpful examples and are initially trained using a ‘golden dataset’, which are available upon request.

Oxford Spec Document Categorization

Summary of Task

- Compare two comments which relate to the same political subreddit.
- Label whether they **agree** with one another, whether they are **neutral** towards each other, or whether the comments **disagree** with one another.
- Select **unsure** if it is not possible to make a decision based on the information at hand.
 - **“Agree”** : the reply shows approval towards the initial statement or initial author
Some examples of phrases that express agreement are: “I agree”, “That’s absolutely right”, “That’s so true”, “Exactly”, “That’s how I feel”, etc.
 - **“Neutral”** : there is a topical exchange between authors but agreement / disagreement is not expressed.
 - **“Disagree”** : the reply shows disapproval or dislike towards the initial statement or initial author.
Some examples of phrases that express disagreement are: “No way”, “I don’t think so”, “Not necessarily”, “That’s not always true”, etc.
 - **“Unsure”** : It is not possible to make a decision based on the information at hand.

I.e. What side do we predict the initial author is on? → Does the reply support this opinion? OR → Do you think the reply author would support this opinion? → If neither of these, select **“Neutral”** or **“Unsure”**

Labels	Attributes
Agree	The reply shows approval towards the initial statement or initial author. Some examples of phrases that express agreement are: “I agree”, “That’s absolutely right”, “That’s so true”, “Exactly”, “That’s how I feel”, etc.
Neutral	There is a topical exchange between authors but agreement/disagreement is not expressed.

Figure 1: Instructions for annotators - first page

Disagree	The reply shows disapproval or dislike towards the initial statement or initial author. Some examples of phrases that express disagreement are: “No way”, “I don’t think so”, “Not necessarily”, “That’s not always true”, etc.
Unsure	It is not possible to make a decision based on the information at hand.

Recommended Workflow:

- Read the Subreddit at the top of the task to understand the **context** of the discussion.
- Read the **“Initial Comment”** to understand the **content and opinion** of the person.
- Read the **“Response”** to determine whether the responder **agrees, disagrees** or is **neutral** towards the **“Initial Comment”**.
- If it is **impossible** to determine whether the responder agrees, disagrees, or is neutral towards the Initial Comment, select **“Unsure”**.

Rules and Tips:

- **Try to understand what each person is saying:**
 - If you don’t understand some of the internet slang being used, google it to find out (e.g based, gaslighted, blue-pilled, red-pilled)
 - Many of these exchanges will be very political (US/UK) so **the more you know about the political people/ideas involved, the better**
 - Check the **“Helpful Abbreviations”** Key at the bottom of each task to explain common Abbreviations used like: **GOP** (Grand Old Party / Republican Party), **POC** (Person of Color), **POTUS** (President of the United States), etc.
 - It’s important to clearly understand what is being said in order to determine if the two comments agree with one another.

Key information to look for:

- **Be on the lookout for keywords in the response that reveal the opinion. This can be key in determining how to label the task.**

Figure 2: Instructions for annotators - second page

- When a clarifying **question is posed** in the **Response**, it can be an indication that the label is **Neutral**
- Look for **Neutral questions** that discuss the same topic but don't reveal the author's stance. This can be a great indicator for a **Neutral** label
 - Here's an example of a **Neutral** Response with a question in it:

Neutral Example	
Initial Comment: I cannot think of anything, but I'd be very interested to hear the post-mortem.	Response: Of the conversation or the country?

- **There can be exceptions to the above rule of thumb:**
 - In addition to being neutral, questions in the **Response** can also be sarcastic or angry, showing **Disagreement**.
 - Keep this in mind when reading a question in the **Response**. Not all questions are **neutral**.
- "Yes!", "I agree", "Exactly" are all great identifier words to show if the **Response agrees** with the **Initial Comment**.
- "No", "I don't think so", "But", "Nah" can show if the **Response disagrees** with the **Initial Comment**.

Don't do this:

- **Do Not** guess the answer when you're unsure.
 - If you're **uncertain** whether the two comments agree or not, or whether they are neutral towards each other, select **"unsure"**.
 - Some comments have nothing to do with one another, for these select **"unsure"** like this example:

Unsure Example	
Initial Comment: C c c c.COMBO BREAKER!!	Response: <u>Ultraaaa Ultraaaaa Ultraaaa</u>

- **Do Not** jump to conclusions
 - If you don't understand some of the words or phrases used in the comments, look it up on google.

Figure 3: Instructions for annotators - third page

A.4 Additional DEBAGREEMENT examples

We provide additional labeled examples from the DEBAGREEMENT dataset, for easy reference.

DISAGREEMENT We provide two examples of disagreement below.

Subreddit: Republican

Submission: Forensic Audit: "We conclude that the Dominion Voting System is intentionally and purposefully designed with inherent errors to create systemic fraud and influence election results."

Parent: So now that they have elected Biden how can we have faith in our election process especially since they didn't listen to the evidence of fraud and chose to ignore and discredit any mention of it.

Child: They haven't, it is a contested election. The media is lying to you, because they ignore the other set.

Subreddit: Brexit

Submission: Why are Brexiteers so angry and furious? I mean brexit has happened, we are at the end of the transition period and all the Anglo headlines are fury at this and fury at that.

Parent: Because the EU has politely declined to fulfill the promises they made?

Child: The EU is fulfilling its promises and helping the UK achieving Brexit

NEUTRAL We provide two examples of neutral interactions below.

Subreddit: Republican

Submission: Nancy Pelosi Let Millions Suffer to Win an Election – Pelosi didn't care about the American people, all she cared about was winning.

Parent: If the people don't vote in GA she's going to keep winning. You have Republicans who are making it easy for her to keep doing what she's doing.

Child: It's because a lot of the Republican base have lost interest in the establishment. They support outsiders like Trump. If the Georgia Senate races aren't helping build Trump's vision for America, what's the point in voting? Republican win or Democrat win it's a lose for these voters.

Subreddit: climate

Submission: Farmers sue German government over missed climate targets: Dismayed by the German government's failure to meet climate protection targets, dairy farmer Heiner Luetke Schwienhorst has filed a lawsuit against Berlin to force it into action.

Parent: The irony of dairy farmers suing over environmental damage is absolutely laughable

Child: Pardon my ignorance, are dairy farmers causing lots of environmental damage?

AGREEMENT We provide two examples of agreement below.

Subreddit: Brexit

Submission: Serious question: where does the idea that the EU will only make a deal at the last minute come from? So I have seen a few times this idea that the EU only break for a deal at the last possible moment, mostly from Leave figures. Where does this idea come from, as seems odd to me when trade deals take many years of negotiating. Are there some examples?

Parent: Misplaced negotiation tactics. - It cannot work when 27 countries have to both understand and agree a "new deal". The UK negotiators seem to think that they are only negotiating with a few suits who have the power to override 27 governments.

Child: Yes. They think EU is only Germany and France. Not 27 independent countries in one Block.

Subreddit: climate

Submission: Democrats to offer \$760B infrastructure plan with big climate theme

Parent: It's always entertaining to watch the Dems wallow over how to pay for their spending bills. You never see that on the other side. How will we pay for invading a new country? With a tax break, of course! They pay for themselves! How do we pay for tax breaks? They pay for themselves! When it comes to infrastructure, something actually proven to increase efficiency, GDP and tax revenue, the Dems navel gaze about how to pay for it.

Child: this is a really interesting point- I'd also add that the right weaponizes this in their claims that we "can't afford it" when discussing any socially beneficial plans.

A.5 Data description

We provide three data files, available and thoroughly described on the dataset download page. The **Raw** file contains all comments from each subreddit prior to the cleaning and filtering steps. The **Labeled Data** file contains all labeled comment-reply interaction pairs whose inter-annotator agreement is at least $\frac{2}{3}$. inter, with labels, filtered to comment interactions that The **Unsure Labels** file contains comment-reply interaction pairs for which the final label is *unsure*, or where annotators have agreement lower than $\frac{2}{3}$.

Labeled data columns The dataset contains the following columns:

label: a column with manually annotated label noting whether the parent and child comment *agree*, *disagree* or have a *neutral* stance towards each other. *0* - disagree; *1* - neutral; *2* - agree. The **Unsure Labels** dataset also contains data with labels of *-100* - unsure.

msg_id_parent: the reddit identifier for the parent message

msg_id_child: the reddit identifier for the child message

submission_id: the reddit identifier for the submission

body_parent: the text of the parent comment

body_child : the text of the child comment

submission_text: abbreviated submission text

subreddit: subreddit name on which the interaction took place

author_parent: reddit username of parent author

exact_time : UTC timestamp of time when interaction occurred

author_child: reddit username of child author

datetime: *datetime* timestamp for when interaction occurred, formatted as YYYY-MM-DD HH:MM:SS

agreement_fraction: the fraction of annotators that agree with the final label

individual_kappa: the Fleiss' kappa for each annotation. Calculation of Fleiss' kappa is taken from appendix A.2 of [8].

A.6 Datasheet for datasets

A.6.1 Motivation

For what purpose was the dataset created? Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description. DEBAGREEMENT was created to better understand the evolution of online debates and to provide annotated data for (dis)agreement detection.

Who created the dataset (e.g., which team, research group) and on behalf of which entity (e.g., company, institution, organization)? Oxford University partnered with Scale AI.

Who funded the creation of the dataset? If there is an associated grant, please provide the name of the grantor and the grant name and number. Data source was free; annotation was provided by Scale AI as part of project collaboration.

Any other comments? No

A.6.2 Composition

What do the instances that comprise the dataset represent (e.g., documents, photos, people, countries)? Are there multiple types of instances (e.g., movies, users, and ratings; people and interactions between them; nodes and edges)? Please provide a description. Instances represent user-to-user online interactions.

How many instances are there in total (of each type, if appropriate)? 42,894 with additional detail provided in Table 1 of the paper.

Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set? If the dataset is a sample, then what is the larger set? Is the sample representative of the larger set (e.g., geographic coverage)? If so, please describe how this representativeness was validated/verified. If it is not representative of the larger set, please describe why not (e.g., to cover a more diverse range of instances, because instances were withheld or unavailable). The dataset is a sample; the sampling technique was developed to ensure meaningful NLP samples were extracted. The procedure is detailed in section 3.

What data does each instance consist of? "Raw" data (e.g., unprocessed text or images) or features? In either case, please provide a description. Each instance contains minimally altered text interactions data, along with the time and username attributes. Example:

Parent: So now that they have elected Biden how can we have faith in our election process especially since they didn't listen to the evidence of fraud and chose to ignore and discredit any mention of it.

Child: They haven't, it is a contested election. The media is lying to you, because they ignore the other set.

Topic: Forensic Audit: "We conclude that the Dominion Voting System is intentionally and purposefully designed with inherent errors to create systemic fraud and influence election results."

Subreddit: Republican

Is there a label or target associated with each instance? If so, please provide a description. Yes, this dataset provides *agree*, *disagree*, or *neutral* labels associated with each interaction.

Is any information missing from individual instances? If so, please provide a description, explaining why this information is missing (e.g., because it was unavailable). This does not include intentionally removed information, but might include, e.g., redacted text. No.

Are relationships between individual instances made explicit (e.g., users' movie ratings, social network links)? If so, please describe how these relationships are made explicit. Yes - via username.

Are there recommended data splits (e.g., training, development/validation, testing)? If so, please provide a description of these splits, explaining the rationale behind them. Yes, described in section 4.2.

Are there any errors, sources of noise, or redundancies in the dataset? If so, please provide a description. No - data with potentially inaccurate annotation labels was removed in pre-processing.

Is the dataset self-contained, or does it link to or otherwise rely on external resources (e.g., websites, tweets, other datasets)? If it links to or relies on external resources, a) are there guarantees that they will exist, and remain constant, over time; b) are there official archival versions of the complete dataset (i.e., including the external resources as they existed at the time the dataset was created); c) are there any restrictions (e.g., licenses, fees) associated with any of the external resources that might apply to a future user? Please provide descriptions of all external resources and any restrictions associated with them, as well as links or other access points, as appropriate. The dataset is self contained, but scraped from Pushshift, which is open source for research purposes, as described in [5].

Does the dataset contain data that might be considered confidential (e.g., data that is protected by legal privilege or by doctor-patient confidentiality, data that includes the content of individuals' non-public communications)? If so, please provide a description. No.

Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety? If so, please describe why. Yes - the data contains minimally altered online interactions (as described in section 3.1) which may, at times, contain hostile and offensive interactions. The data is purposefully not censored to accurately capture the nature of polarisation and online debates.

Does the dataset relate to people? If not, you may skip the remaining questions in this section. Yes.

Does the dataset identify any subpopulations (e.g., by age, gender)? If so, please describe how these subpopulations are identified and provide a description of their respective distributions within the dataset. No.

Is it possible to identify individuals (i.e., one or more natural persons), either directly or indirectly (i.e., in combination with other data) from the dataset? If so, please describe how. No.

Does the dataset contain data that might be considered sensitive in any way (e.g., data that reveals racial or ethnic origins, sexual orientations, religious beliefs, political opinions or union memberships, or locations; financial or health data; biometric or genetic data; forms of government identification, such as social security numbers; criminal history)? If so, please provide a description. No.

Any other comments? No.

A.6.3 Collection Process

How was the data associated with each instance acquired? Was the data directly observable (e.g., raw text, movie ratings), reported by subjects (e.g., survey responses), or indirectly inferred/derived from other data (e.g., part-of-speech tags, model-based guesses for age or language)? If data was reported by subjects or indirectly inferred/derived from other data, was the data validated/verified? If so, please describe how. Data collection is detailed in section 3; raw data is available in the dataset drive.

What mechanisms or procedures were used to collect the data (e.g., hardware apparatus or sensor, manual human curation, software program, software API)? How were these mechanisms or procedures validated? Data collection is detailed in section 3.

If the dataset is a sample from a larger set, what was the sampling strategy (e.g., deterministic, probabilistic with specific sampling probabilities)? Data sampling is detailed in section 3.

Who was involved in the data collection process (e.g., students, crowdworkers, contractors) and how were they compensated (e.g., how much were crowdworkers paid)? The authors of this paper collected the data; they were compensated with their PhD stipends.

Over what timeframe was the data collected? Does this timeframe match the creation timeframe of the data associated with the instances (e.g., recent crawl of old news articles)? If not, please describe the timeframe in which the data associated with the instances was created. The data was created over several years; it is stored in a data repository, allowing it to be collected over the course of several days.

Were any ethical review processes conducted (e.g., by an institutional review board)? If so, please provide a description of these review processes, including the outcomes, as well as a link or other access point to any supporting documentation. No.

Does the dataset relate to people? If not, you may skip the remaining questions in this section. Yes.

Did you collect the data from the individuals in question directly, or obtain it via third parties or other sources (e.g., websites)? Third parties - websites.

Were the individuals in question notified about the data collection? If so, please describe (or show with screenshots or other information) how notice was provided, and provide a link or other access point to, or otherwise reproduce, the exact language of the notification itself. No - the publication of the data is part of the agreement of using reddit.

Did the individuals in question consent to the collection and use of their data? If so, please describe (or show with screenshots or other information) how consent was requested and provided, and provide a link or other access point to, or otherwise reproduce, the exact language to which the individuals consented. Yes - this is part of the agreement of using reddit.

If consent was obtained, were the consenting individuals provided with a mechanism to revoke their consent in the future or for certain uses? If so, please provide a description, as well as a link or other access point to the mechanism (if appropriate). Yes.

Has an analysis of the potential impact of the dataset and its use on data subjects (e.g., a data protection impact analysis) been conducted? If so, please provide a description of this analysis, including the outcomes, as well as a link or other access point to any supporting documentation. No.

Any other comments? No.

A.6.4 Preprocessing/cleaning/labeling

Was any preprocessing/cleaning/labeling of the data done (e.g., discretization or bucketing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances, processing of missing values)? If so, please provide a description. If not, you may skip the remainder of the questions in this section. Yes - details are provided in section 3.

Was the “raw” data saved in addition to the preprocessed/cleaned/labeled data (e.g., to support unanticipated future uses)? If so, please provide a link or other access point to the “raw” data. Yes - raw data is provided in the data repository.

Is the software used to preprocess/clean/label the instances available? If so, please provide a link or other access point. Yes - this is publicly available Python software.

Any other comments? No.

A.6.5 Uses

Has the dataset been used for any tasks already? If so, please provide a description. Yes, it was used for the analysis throughout the paper.

Is there a repository that links to any or all papers or systems that use the dataset? If so, please provide a link or other access point. No.

What (other) tasks could the dataset be used for? This point is discussed within section 4.

Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses? For example, is there anything that a future user might need to know to avoid uses that could result in unfair treatment of individ-

uals or groups (e.g., stereotyping, quality of service issues) or other undesirable harms (e.g., financial harms, legal risks) If so, please provide a description. Is there anything a future user could do to mitigate these undesirable harms? Nothing beyond what is listed in section 3.

Are there tasks for which the dataset should not be used? If so, please provide a description.
No.

Any other comments? No.

A.6.6 Distribution

Will the dataset be distributed to third parties outside of the entity (e.g., company, institution, organization) on behalf of which the dataset was created? If so, please provide a description.
No

How will the dataset will be distributed (e.g., tarball on website, API, GitHub)? Does the dataset have a digital object identifier (DOI)? The dataset is distributed via Google Drive for the reviewers and will be distributed via Scale AI website after NeurIPS reviews.

When will the dataset be distributed? The dataset is available to the reviewers, and will be available publicly at the end of the review process.

Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)? If so, please describe this license and/or ToU, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms or ToU, as well as any fees associated with these restrictions. Dataset Terms of Use: Scale AI, Inc. and the University of Oxford and their affiliates strive to enhance public access to and use of data that they collect, annotate, and publish. The data are organized in datasets listed at <https://scale.com/open-datasets/oxford>. The Datasets are collections of data, managed by Licensors and provided in a number of machine-readable formats. Licensors provide any individual or entity (hereinafter “You” or “Your”) with access to the Datasets free of charge subject to the terms of this agreement (hereinafter “Dataset Terms”). Use of any data derived from the Datasets, which may appear in any format such as tables, charts, devkit, documentation, or code, is also subject to these Dataset Terms. By downloading any Datasets or using any Datasets, You are agreeing to be bound by the Dataset Terms. If you are downloading any Datasets or using any Datasets for an organization, you are agreeing to these Dataset Terms on behalf of that organization. If you do not have the right to agree to these Dataset Terms, do not download or use the Datasets.

Licenses: Unless specifically labeled otherwise, these Datasets are provided to You under a Creative Commons Attribution 4.0 International Public License (“CC BY 4.0”), with the additional terms included in these Dataset Terms. The CC BY 4.0 may be accessed at <https://creativecommons.org/licenses/by/4.0/legalcode>. When You download or use the Datasets from the Website or elsewhere, You are agreeing to comply with the terms of CC BY 4.0. Where these Dataset Terms conflict with the terms of CC BY 4.0, these Dataset Terms will control.

Have any third parties imposed IP-based or other restrictions on the data associated with the instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms, as well as any fees associated with these restrictions. No.

Do any export controls or other regulatory restrictions apply to the dataset or to individual instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any supporting documentation. No

Any other comments? No.

A.6.7 Maintenance

Who will be supporting/hosting/maintaining the dataset? The dataset will be made available both by the team from Scale AI and the Oxford authors. Their emails are provided on the title page.

How can the owner/curator/manager of the dataset be contacted (e.g., email address)? The authors are responsible for this - contact details are provided on the title page.

Is there an erratum? If so, please provide a link or other access point. No.

Will the dataset be updated (e.g., to correct labeling errors, add new instances, delete instances)? If so, please describe how often, by whom, and how updates will be communicated to users (e.g., mailing list, GitHub)? No

If the dataset relates to people, are there applicable limits on the retention of the data associated with the instances (e.g., were individuals in question told that their data would be retained for a fixed period of time and then deleted)? If so, please describe these limits and explain how they will be enforced. No.

Will older versions of the dataset continue to be supported/hosted/maintained? If so, please describe how. If not, please describe how its obsolescence will be communicated to users. Not applicable.

If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so? If so, please provide a description. Will these contributions be validated/verified? If so, please describe how. If not, why not? Is there a process for communicating/distributing these contributions to other users? If so, please provide a description. Not applicable.

Any other comments? No.

A.7 Author statement:

The authors confirm that they bear all responsibility in case of violation of rights.