Prompt, Condition, and Generate
Classification of Unsupported Claims with In-Context Learning
Christensen, Peter Ebert; Yadav, Srishti; Belongie, Serge

Publication date:
2023

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY

Citation for published version (APA):
Abstract

Unsupported and unfalsifiable claims we encounter in our daily lives can influence our view of the world. Characterizing, summarizing, and – more generally – making sense of such claims, however, can be challenging. In this work, we focus on fine-grained debate topics and formulate a new task of distilling, from such claims, a countable set of narratives. We present a crowdsourced dataset of 12 controversial topics, comprising more than 120k arguments, claims, and comments from heterogeneous sources, each annotated with a narrative label. We further investigate how large language models (LLMs) can be used to synthesize claims using In-Context Learning. We find that generated claims with supported evidence can be used to improve the performance of narrative classification models and, additionally, that the same model can infer the stance and aspect using a few training examples. Such a model can be useful in applications which rely on narratives, e.g., fact-checking.

1 Introduction

Online platforms have revolutionized the landscape of public discourse, facilitating extensive debates across a wide range of topics. However, these online discussions often suffer from a lack of coherent and concise arguments. Despite this inherent challenge, it is possible to discern particular motions (Levy et al., 2014), opinions (Li et al., 2020), human values (Kiesel et al., 2022), and narratives (Christensen et al., 2022) within the seemingly disorganized discourse. The ability to identify narratives in online debates is paramount for fact-checking and argument mining, as it enables the evaluation of unsupported claims and their validity.

In our methodology, narratives are differentiated from arguments and claims by incorporating additional attributes: topics, stances, aspects, and evidence. The topic refers to the subject under discussion, such as the ethical aspects of cloning humans for reproductive purposes. The stance represents the viewpoint taken on the topic, for example, a negative stance indicating that cloning for reproductive purposes is considered unethical and unacceptable. Within the narrative, the aspect focuses on a specific perspective, providing a more nuanced understanding of the topic. For instance, the aspect within the context of cloning could be the creation of cloned embryos solely for research purposes, which delves into a particular subtopic. These attributes aid in identifying arguments in non-argumentative sources (Stab et al., 2018).

Evidence plays a crucial role in assessing the credibility of a statement. When supported by evidence, a statement gains strength and credibility, classifying it as an argument (Hansen and Hershcovich, 2022). For instance, in the text “Cloning humans for reproductive purposes is unethical and unacceptable, but creating cloned embryos solely for research – which involves destroying them anyway – is downright criminal,” the presence of evidence highlighting the destruction of embryos strengthens the argument. Conversely, without evidence, a statement such as “Cloning humans for reproductive purposes is unethical” would be categorized as a claim, lacking the necessary substantiation to be considered an argument. Additional support is required to validate a claim as an argument. With these definitions, we can briefly differentiate between narrative, arguments and claims as:

1. Narrative: Concise expression of an individual’s perspective on a specific topic.
2. Claim: Statement or proposition without supporting evidence.
3. Argument: Claim supported by evidence and reasoning. Aim to justify a specific stance on a topic

As seen above, claims can lack or have insufficient evidence and be unverifiable or unfalsifiable for purposes of fact-checking in real world scenarios (Glockner et al., 2022) and are hence often not suitable for fact-checking pipelines and
thus discarded (Augenstein, 2021). Instead of discarding the claims and arguments, we propose that one should instead identify the individual unsupported claim or narrative, e.g., “human cloning is wrong”. We call this task Narrative Prediction which forms the basis of this paper. We use the word “Prediction” as an umbrella term, as it can be viewed as either classification (due to the small set of unsupported claims that reflect different viewpoints in the debate) or alternatively a generation task. In addition to fact-checking the existing literature on claim generation using large language models (LLMs) lacks attention to the relationship between narratives extracted from online debate portals (Christensen et al., 2022) and argumentative texts (Habernal and Gurevych, 2016), as well as the effective modeling of narratives by LLMs. This work addresses these gaps by formalizing narratives in online debates and understanding their elements: topics, aspects, stance, and evidence. Additionally, a curated dataset of 120k tweets, with around 40 narratives per topic, is introduced to train and evaluate narrative prediction techniques for fact-checking systems. Furthermore, we propose a method to enhance narrative prediction by generating synthetic tweets through argumentative attributes such as stance and aspects using few-shot In Context Learning (ICL) as illustrated this in Fig. 1. The task of narrative prediction simply corresponds to only the right hand side of the figure using no generated candidates where we fine-tune the LLM on tweets. In summary, the contributions of this paper are:

1. A specific definition for narratives, along with an analysis of how this differs from arguments, claims, and motions.
2. A new dataset and task, consisting of online comments and tweets labelled for narrative prediction.
3. A narrative prediction approach that maps all the tweets from a fine-grained debate into a list of narratives using a LLM.
4. A computational approach that generates synthetic arguments/claims with a specified aspect and stance.

2 Related Works

Corpora of textual claims considering various controversial topics have often been used in the study of rhetoric and argumentation, including summarization (Stammbach and Ash, 2020), optimization (Skitalinskaya et al., 2022), identifying human values (Kiesel et al., 2022), robustness of arguments (Sofi et al., 2022), controllable text generation (Schiller et al., 2021), stance detection (Stab et al., 2018), and studying what constitutes an argument (Trautmann et al., 2020). Prior work on claim and argument summarization has been beneficial in different tasks and domains. In early works, summarization was used for explainable fact-checking (Stammbach and Ash, 2020; Mishra et al., 2020) and has recently been used to denoise tweets (Bhatnagar et al., 2022). However, abstractive summarization techniques for real-world tweets are still underdeveloped compared to traditional text summarization methods. Given the effectiveness of prompt-based methods in tasks like abstractive summarization and binary classification (Chung et al., 2022; Sanh et al., 2022), we propose exploring these methods to enhance the text generation of arguments, particularly within the fine-grained topic debates. While fine-grained approaches have been explored in argument mining (Hansen and Hershcovich, 2022; Trautmann et al., 2020; Schiller et al., 2021), they often address broader controversial topics (“minimum wage.”) rather than narrow debates (“crypto currencies as a fiat currency.”). Similarly, other works that classify if a claim is mentioned in a text (Mirkin et al., 2018) studies motions which are an action that should be taken (as can be seen with the example “we should introduce goal line technology”). Other lines of works focus on scaling up by detecting “generic” claims frequent across topics (Orbach et al., 2019) or mining candidate claims from corpora (Lavee et al., 2019). In comparison we are envisioning our work to be applicable for unfalsifiable or unverifiable claims coming from short noisy tweets rather than a high quality curated database (iDebate) which contain minutes long speeches.

In our work we create a new dataset, focusing on narrow debate topics, by relying on an argument mining annotation scheme based on Hansen and Hershcovich (2022), consisting of various categories of claims and arguments found in online debates. Where Hansen and Hershcovich (2022) compare arguments in terms of categories (normative or factual arguments), we propose and study the new task of predicting controversial narratives from tweets. Perhaps most similar to our work is Christensen et al. (2022), which proposed a human-in-the-loop-based model to cluster unfalsifiable claims using crowdsourced triplets similarities.
Figure 1: Prompt, Condition, and Generate: A framework to enhance narrative prediction by synthetizing tweets. We first prompt a LLM for the stance and aspects of a new tweet using ICL with some examples, we then condition the LLM on these attributes to synthesize tweets. Lastly, we fine-tune a LLM on all tweets to generate narratives.

3 Task and Data

This section introduces our definition of a narrative, and a proposed task, and presents the data used for development and evaluation.

3.1 Narrative Definition

As mentioned in the introduction, we define the term narrative as a concise statement lacking supporting evidence, which can originate from an unfalsifiable or unverifiable claim. Additionally, narratives can include arguments supported by evidence types such as anecdotal, expert, or normative sources as defined in (Hansen and Hershcovich, 2022), instead of empirical studies. The objective of our methodology is to identify a small set of unsupported claims that reflect diverse viewpoints in a debate and require attention from fact-checkers.

After defining the term narrative we can now focus on a theoretical underpinning of this paper which is a proposed relationship between number of narratives and the scope of the fine-grained debate, that we call the parrot hypothesis.

The Parrot Hypothesis  

In a given social media debate, the thoughts and opinions contributed by commentators resolve to a finite set of distinct narratives. While users could, in principle, state their views in a concise, distilled manner, they often prefer to write embellished variants or personal takes that require reading between the lines.

At its core, the parrot hypothesis seeks to propose a concept to manage the variations of statements in a debate. By grouping statements into a finite set of narratives related to common topics, the hypothesis narrows the scope of the debate and transforms it into a classification problem. Narratives, representing individual unsupported claims or viewpoints, play a crucial role in capturing diverse perspectives and supporting fact-checking efforts. Despite the potential for infinite arguments, a limited number of distinct claims tend to emerge in online debates, backed by the majority of users (Boltužić and Šnajder, 2015). Our hypothesis is that a narrow enough topic will emit such behaviour from users. Incorporating the parrot hypothesis and identifying narratives could enable a more systematic analysis to improve understanding of narratives and facilitating fact-checking.

3.2 Narrative Prediction

We approach the problem of narrative prediction on social media, specifically focusing on tweets.

**Task**  
Given a single tweet $t$, a statement by a participant in a debate, and a set of possible narratives $\mathcal{N}$, rewrite $t$ into a narrative $n \in \mathcal{N}$ such that:

- the narrative is written as an unsupported claim,
- only one narrative $n$ can be selected for each tweet from $\mathcal{N}$, and
- $n$ preserves the meaning of $t$ as much as possible.

The set of possible narratives, denoted as $\mathcal{N}$, is sourced from domain experts. Although a tweet may implicitly or explicitly contain multiple nar-
narratives, our aim is to identify and assign only one explicitly stated narrative for each tweet.

By addressing narrative prediction in this manner, we strive to transform tweets into coherent and explicit unsupported claims, contributing to a deeper understanding and analysis of the content within the context of social media discourse.

3.3 Annotation scheme

To collect relevant data, we use an annotation scheme comprising a fine-grained topic, a sentence, and a narrative (unfalsifiable and unverifiable claim). Additionally, we explore the augmentation of an existing dataset, following an alternative annotation scheme (Schiller et al., 2021), to incorporate attributes such as stance (polarity of the argument) and aspect (subtopics or viewpoints) and use it for the generation of synthetic tweets. Though there exist narratives that are claims with supporting evidence (can be anecdotal, factual or normative which are found in (Hansen and Hershcovich, 2022)), the type of evidence is not considered for annotation. The details of this will be explained in the following section.

3.4 Dataset creation

We present two datasets: Twitter-Narratives-9 (TN9) and an augmented version of the UKP-Corpus-Aug dataset. UKP-Corpus-Aug, which is the augmented UKP-Corpus dataset includes stance, aspect, and narrative annotations for three randomly selected topics from the original UKP-Corpus (Schiller et al., 2021; Stab et al., 2018). On the other hand, TN9 consists of narrative annotations for 9 carefully selected controversial topics. Table 1 provides an overview of the datasets, including key statistics and a comparison between them. Additionally, Table 2 presents examples from TN9.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Sentence</th>
<th>Narrative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crypto</td>
<td>you are promoting crypto which is a scam helping thieves and criminals</td>
<td>Influencers are scamming their fans using crypto</td>
</tr>
<tr>
<td></td>
<td>you are also full of plastic parts and fillers profitable for the</td>
<td></td>
</tr>
<tr>
<td></td>
<td>pharmaceutical and cosmetic industry</td>
<td></td>
</tr>
<tr>
<td>Formula</td>
<td>My congressman here voted NO on, lowering gas prices, NO on the baby</td>
<td>People are reselling baby formula to other countries</td>
</tr>
<tr>
<td></td>
<td>formula bill, NO on contraception (17%), and NO on other helpful bills</td>
<td>for higher prices</td>
</tr>
<tr>
<td></td>
<td>It is unbecoming to complain about economic hardship and then contribute to it.</td>
<td></td>
</tr>
<tr>
<td>AGI</td>
<td>And on the other side, AGI will be the single greatest technology to</td>
<td>AI will not replace humans but augment them</td>
</tr>
<tr>
<td></td>
<td>alleviate human suffering in all of history</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Example sentences and annotated narratives.

3.4.1 Scraping

We start with scraping relevant data from Twitter. First a series of searches is executed combining different keywords and sentences/phrases, highlighting different statements in a topic. We search for 40 different keywords per topic from year 2016-2022 and search for as many fields (e.g., images, links, and other metadata) as possible using the Twitter API. The specific keywords used for each topic can be found in Appendix B.

3.4.2 Filtering and Data Cleaning

To ensure that we are working with claims, we perform filtering steps. First, we remove duplicates but maintain identical sentences with different hashtags after removing retweets, quote tweets, links and videos, as well as mentions of users, token and media mentions. Second, we replace unreadable hexadecimal representations of unicode characters with their respective character, and encode the text with ascii characters. This results in 98,187 English tweets in total, around 11k tweets for each topic. The geographic distribution of the tweets as shown in Figure 2.

3.4.3 Dataset Annotation

Annotation of narratives is conducted using Amazon Mechanical Turk in 2 rounds. In round 1, we design a pretest to ensure that the workers know the difference between an argument, claim and evidence using the dataset from (Hansen and Hershcovich, 2022). Given a tweet the annotators classify whether the tweet is a claim, argument or neither. Furthermore they also classify the evidence type given an argument (Hansen and Hershcovich, 2022). Learning to distinguish between these claims will help them in determining the narrative of the tweet. After passing at least 4 out of 5 questions the workers could begin annotating our 120k tweets. Using lists of around 40 narratives
per topic compiled by domain experts we proceed to round 2. Each task consists of 1 tweet from 1 topic and annotators are asked to pick 1 out of ~ 40 different narratives this tweet follows (if any). Furthermore, by using the definition of argument as in Hansen and Hershcovich (2022) we can decompose it into 1) a claim and 2) its evidence, and use their schema to categories the type of evidence an argument may have. We thus additionally ask if the tweet is a claim or an argument with an evidence type that is not a study. A study here refers to “Results of a quantitative analysis of data, given as numbers, or as conclusions.” That is statements that are cited, or easy to verify numbers should they appear in another argument. (Hansen and Hershcovich, 2022) The pay is 18$/hr. Detailed instructions can be found in Appendix D.

Figure 2: Visualization of the percentages of the number of tweets per country. Like (Huang and Carley, 2019) only 2% of all tweets had available geotags and the tweets are found to be predominantly from the US, where the userbase is numerically the largest.

4 Method

Problem statement: Our goal is to create a model that output an estimate of the true narrative \( n \) given tweet \( t \) from debate \( d \). We do so by

- Investigating if identifying the narrative of a tweet is best suited as a text2text approach or a classification approach.
- Creating a data augmentation step using different kinds of ICL to help the best fine tuning procedure.

Narrative prediction approach Given observed data \( \{t_i, n_i\}_{i=1}^N \), we could parameterize a model as \( \hat{n} = f_{\phi}(t) \), where \( f \) is a pretrained LLM with parameters \( \phi \) is the model that we finetune. This is illustrated as step 3 in fig. 1. We note that our prediction \( \hat{n} \) in this case would be free-form text, a text2text approach. As the last step in fig. 1, illustrates a finetuning procedure we could alternatively parameterise a model as \( \hat{n} = g(h_{\theta}(f_{\omega}(t))) \), where \( g \) is a lookup function that maps class \( c_i \) to narrative \( n_i \) for debate \( d \), \( h \) is a multi-class classifier with parameters \( \theta \), and \( f \) the model from before, but only using the encoder with parameters \( \omega \) and the classification head \( h \). The prediction \( \hat{n} \) is now a class, a multi-class classification approach. The lookup function \( g \) enables us to take a predicted class and look up the actual narrative in the list written in appendix F. But doing this substitution we can calculate a Rouge score between the target narrative and the predicted one. However, during training we opt for optimizing the crossentropy loss on the narrative classes.

Prompt, Condition and Generate In addition to only finetuning \( f \), which is a LLM, we argue for using new methodologies focusing on ICL to exploit the capabilities of the LLM further and validate their performance on our new dataset. As such we let us inspire by prior work Schiller et al. (2021) on generating synthetic arguments (candidates) using aspects and a stance, that we call Prompt, Condition and Generate (PCG) and add candidates to our finetuning procedure as illustrated in step 3 in fig. 1. In contrast to their work our setup requires no training and can be done using a few examples using ICL as illustrated in fig. 1. That is as a first step we annotate a few handmade examples with topics names, a binary stance and a snippet of the example tweet which forms the aspect and then we prompt the frozen model to output the stance and aspect of a novel tweet \( t \). Then we condition the same model anew on its predicted stance and aspect to generate a candidate, by asking it to write a tweet knowing only about the debate topic, a stance and the aspect. To complete the creation of synthetic data we copy the original narrative \( n \) from \( t \) to form the candidate. This data can be used for additional fine-tuning of the text generation model that generate narra-tives, as shown in fig. 1.

ICL Methods In the context of using LLM for ICL, a simple but effective approach called few-shot learning is to provide several examples of a task in the same prompt with the given input (Brown et al., 2020). Additionally one can also first generate an explanation as to why certain outputs are favourable before generating the final answer, this is called Chain-of-Thought (CoT) (Wei et al., 2023). Furthermore one should be careful with the selected examples for ICL. As noted in Zhao et al. (2021), standard ICL can be biased towards
the training examples and the order of their occurrence. To mitigate this effect one can estimate the bias towards each answer by feeding in a test input that is content-free, e.g., “N/A” and “”. In practice one can fit an affine transformation to “calibrate” (Cal) the model’s output probabilities to cause a uniform prediction for “N/A”. We will investigate these methods in our PCG setup using different numbers of shots for aspect and stance prediction.

5 Experiments

In this section we investigate the performance of our finetuning approaches, including synthetic tweet generation for performance enhancement and the subtasks of stance and aspect prediction using different ICL techniques. We predict narratives on 7548 test cases (629 per topic).

5.1 Setup

Classification: As described in Section 4.1, our classification model ($SFT_{head}$) consists of an encoder $f_\omega$, being a T0 encoder (Sanh et al., 2022) model and a multi-class classifier $h$, which is a single MLP that project the hidden layer down to the number of narratives present in one topic. We only finetune $h$ using using the crossentropy loss. Finally using $g$ we can report the Rouge-L F1 score as we convert the predicted class into the written narrative and comparing it with ground truth.

Text generation: In contrast to the classification model $f_\omega$ is the full T0 model. We add new parameters and make a parameter efficient fine-tuning setup known as LoRA (Liu et al., 2022) on the T0 model. LoRA incorporates two low-rank matrices that are added to each parameter matrix in T0. We measure the Rougle-L F1 score between the generated narrative $\hat{n}$ and the ground truth narrative.

Prompt, Condition and Generate: To enhance the above mentioned setups we generate synthetic tweets. We do this we first infer the stance and aspect of a new tweet by insert up to 4 such examples into the prompt, and then second simply prompt our frozen model to write a tweet with the predicted stance containing the sentence from the aspect and about the same topic as the new tweet, this is shown in Figure 1.

In addition to generating candidates with our original model $f$, we test the generality of generating candidates of these attributes by conditioning other LLMs on them. These include T5-Flan-3B (Chung et al., 2022), BLOOM-175B (Workshop et al., 2023) and CTRLUKP (Schiller et al., 2021) for the UKP-Corpus. When finetuning using candidates and tweets, we compare them using several metrics like precision oriented BLEU (Papineni et al., 2002), recall oriented Rouge-L (Lin, 2004), METEOR (Banerjee and Lavie, 2005), and finally chrF (Popović, 2015). To automatically quantify to what extent a candidate contains the meaning of the original claim, we compute their semantic similarity in each case using the BERT-score(Zhang et al., 2020). Additionally we conduct a human evaluation of the generated candidates to ensure readability for humans. For each generative model and topic we select 10 candidates and acquire 2 independent crowdworkers via MTurk at 18$/hour. The annotators scored all candidates on four quality metrics: (1) argument quality (2) persuasiveness, (3) meaning preservation and (4) fluency. We follow Schiller et al. (2021) for assessing the argument quality, Habernal and Gurevych (2016) for persuasiveness and Skitalinskaya et al. (2022) for quality using these Likert scales: Argument Quality. 1 (much worse than original) - 5 (notably improved), Persuasiveness. 1 (generated text less persuasive than original) - 3 (generated text is more persuasive), Fluency. 1 (major errors) - 3 (fluential) and Meaning Preservation. 1 (entirely different) - 5 (identical). Lastly we report the inter-annotator
agreement (Cohen, 1960) and krippendorff’s alpha (Krippendorff, 2004) between 2 annotators.

Stance prediction  To do stance prediction, we classify a tweet as either “for” or “against” a particular topic. For ICL methods we output the most likely word and convert it to 0 or 1 to compare it with the binary class output for the baseline model. We use binary cross entropy loss to compare the predicted stance with the true label.

Aspect prediction  Here we aim to identify the correct span of text within a tweet. The span is represented using the beginning-inside-outside (BIO) tags format (Ramshaw and Marcus, 1995). Here the initial word within the span is given the label “B” for beginning, the following words within the span is given the label “I” for inside, and finally any other word is given the label “O” for outside, making it a ternary classification task for the baselines. We sample multiple completions using beam search and report the average micro F1 and accuracy for both stance and aspect prediction.

5.2 Results

Classification  10 epochs of fine-tuning LM head results in a 38.54 Rouge-L F1 score for the UKP corpus and 38.72 Rouge-L F1 score for TN9.

Text generation  Fine-tuning the LoRA weights results in a 39.32 Rouge-L F1 score for the UKP corpus and 39.49 Rouge-L F1 score for TN9, similar to other summarization tasks (Zhang et al., 2022). Inspecting the results of these models for a couple of outputs is shown in Table 3. Analysing the second example we see a more concisely written narrative than the target, this lowers the resulting Rouge F1 score due to its shorter common sub-sequence.

Table 3: Sentences with predicted and target narrative.

<table>
<thead>
<tr>
<th>Tweet t</th>
<th>Model Prediction t₀</th>
<th>Target Narrative t₁</th>
</tr>
</thead>
<tbody>
<tr>
<td>Animals are not ingredients</td>
<td>Eating meat is murder</td>
<td>Eating meat is murder</td>
</tr>
<tr>
<td>Yall find hypermasculinity resulting in insecurities about the lack of a better body attractive? Lmaaoo</td>
<td>Hypermasculinity is problematic</td>
<td>Hypermasculinity in and of itself is the problem</td>
</tr>
</tbody>
</table>

Table 5 shows the quantitative metrics of our candidates. The relatively low BLEU (6.5) and ROUGE-L (9.2) indicate that revisions take place, however due to the high BERT-score (90.5) the meaning is largely preserved. Also the METEOR and Rouge-L scores are similar to Schiller et al. (2021) indicating similar generative behaviour. TN9 lower scores indicate that the model has difficulty generating similar sentences to the original tweet using a predicted stance and aspect.

Table 4: Summary of Rouge-L F1 scores using text2text supervised fine-tuning on the original dataset as well as candidates generated with different models.

Prompt condition generate  Given prior results we proceed with the best setup, the text generation setup from before and Table 4 shows the average Rouge-L F1 micro accuracy using additional candidate examples generated by T0-3B, T5-flan-3B model (Chung et al., 2022), an API call to BLOOM-176B (Workshop et al., 2023) and the CTRL generative model from Schiller et al. (2021) respectively. Using 629 candidates we get a 4 percentage point increase from the 39.49 Rouge-L F1 for TN9 from before, highlighting the strength of our approach.

Table 5: Automatic evaluation: Average performance of each model on 629 test cases per topic

Table 6 shows generally low Krippendorff’s alpha agreement of 0.24 on average, which are common in subjective tasks (Wachsmuth et al., 2017). The inter-annotator agreement (Cohen, 1960) varies from model and attribute but is on average .25, which can be interpreted as “fair” agreement (Landis and Koch, 1977). Table 6 shows that human annotators find text generated by T0, having a higher persuasiveness (2.6) and having similar meaning to the source text (4.5) than the other methods. However, candidates from BLOOM and CTRL-UKP have a higher argument quality (3.5 vs. 3.6 and 4.2) and are more fluently written. Table 7 shows T0 being preferred for generating meaningful and persuasive texts. This is important as we will use the data in a fine-tuning setup.
Figure 3: Average Aspect accuracy of few-shot ICL (T0−3B) on the Abortion, Cloning and the Nuclear Energy topic in the UKP dataset using random subsets of \( k' = 1 \ldots 4 \) examples. We display the best performances of the best fine-tuned BERT\textsubscript{BASE} baselines, the tags \textit{only}, \textit{remain} and \textit{all} indicate the same setup from table 8.

<table>
<thead>
<tr>
<th>Model</th>
<th>Persuasiveness</th>
<th>Fluency</th>
<th>Argument</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTRLUKP</td>
<td>2.1</td>
<td>2.3</td>
<td>3.6</td>
<td>3.4</td>
</tr>
<tr>
<td>BLOOM</td>
<td>1.9</td>
<td>2.8</td>
<td>4.2</td>
<td>4.1</td>
</tr>
<tr>
<td>T5-flan</td>
<td>2.2</td>
<td>1.8</td>
<td>3.2</td>
<td>3</td>
</tr>
<tr>
<td>T0</td>
<td>2.6</td>
<td>2.8</td>
<td>3.5</td>
<td>4.5</td>
</tr>
<tr>
<td>TN9</td>
<td>2</td>
<td>2.7</td>
<td>3.4</td>
<td>3.5</td>
</tr>
<tr>
<td>T5-flan</td>
<td>2.4</td>
<td>2.3</td>
<td>3.6</td>
<td>3.3</td>
</tr>
<tr>
<td>T0</td>
<td>2.4</td>
<td>2.5</td>
<td>3.4</td>
<td>3.5</td>
</tr>
</tbody>
</table>

Table 6: Human evaluation: Average scores on 10 candidates per topic using different models.

<table>
<thead>
<tr>
<th>Model</th>
<th>Persuasiveness</th>
<th>Fluency</th>
<th>Argument</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTRLUKP</td>
<td>0.2/0.3</td>
<td>0.2/0.3</td>
<td>0.2/0.2</td>
<td>0.4/0.4</td>
</tr>
<tr>
<td>BLOOM</td>
<td>-0.1/-0.1</td>
<td>0.4/0.4</td>
<td>0.1/0.2</td>
<td>0.3/0.1</td>
</tr>
<tr>
<td>T5-flan</td>
<td>0.1/0.1</td>
<td>0.3/0.3</td>
<td>0.1/0.4</td>
<td>0.5/0.4</td>
</tr>
<tr>
<td>T0</td>
<td>0.3/0.3</td>
<td>0.4/0.3</td>
<td>0.2/0.3</td>
<td>0.5/0.3</td>
</tr>
</tbody>
</table>

Table 7: Annotator agreement (Cohens kappa and krippendorffs alpha) using 2 annotators across all topics.

Table 8: Average micro F1 and accuracy for stance prediction using BIO tags.

<table>
<thead>
<tr>
<th>Method</th>
<th>Abor. (F1 / Acc)</th>
<th>Clon. (F1 / Acc)</th>
<th>Nucl. (F1 / Acc)</th>
</tr>
</thead>
<tbody>
<tr>
<td>only</td>
<td>50.1 / 53.1</td>
<td>75.5 / 75.8</td>
<td>37.1 / 58.9</td>
</tr>
<tr>
<td>ICL</td>
<td>54.4 / 53.8</td>
<td>59.3 / 54.9</td>
<td>54.9 / 52.9</td>
</tr>
<tr>
<td>CoT</td>
<td>55.7 / 54.7</td>
<td>62.4 / 56.7</td>
<td>57.4 / 53.6</td>
</tr>
<tr>
<td>Cal</td>
<td>57.3 / 55.6</td>
<td>60.6 / 55.9</td>
<td>58.7 / 54.1</td>
</tr>
<tr>
<td>remain</td>
<td>36.1 / 56.4</td>
<td>35.6 / 55.3</td>
<td>37.1 / 58.9</td>
</tr>
<tr>
<td>all</td>
<td>52.6 / 55.6</td>
<td>77.1 / 77.6</td>
<td>37.1 / 58.9</td>
</tr>
</tbody>
</table>

Table 9: Average micro F1 and accuracy for aspect prediction using BIO tags.

6 Conclusion

In this paper we introduced a new definition of narratives and how to model these in fine-grained debates with large language models. Our approach is based on parameter efficient fine-tuning using controlled text generation using attributes predicted using a handful of examples. We show that claims generated using our approach are genuine and sensible in general. We fine-tune of model on our own dataset and the augmented UKP-corpus and outperform baseline approaches. In future work, we seek to examine multiple completions and ensembles similar to (Liévin et al., 2023) which enables to include examples of up to 100 examples for ICL, to reduce variance and outperform single-sample CoT methods using larger models (GPT-4, ChatGPT, LLama). Moreover, our approach considers each topic independently using a LLM but could be made to consider all simultaneously.
Acknowledgements
PEC, SY and SB was supported by the Pioneer Centre for AI, DNRF grant number P1.

7 Limitations

Scaling to multiple topics For our approach, the prediction of narratives is topic specific and the number of models scales linearly with the with the topics. This is primarily because both the baseline method using a LM head cannot predict new classes and for the text2text approach it is theoretically possible to simply use one model, though initial experiments suggested a model per topic worked better. Instead of directly predicting the narratives, one could instead have ranked the list of narratives given a tweet. This gives us contextual information about the narratives, since they are written in text and not just as a class and provides a number of benefits including having one model for all topics but also new topics. Additionally it could also provide temporal evaluations by adding new emerging narratives to the list.

Scaling to more narratives The current approach requires a domain expert to writing down the particular narratives from the fine grained debate and does not model that there is a countable number of narratives within a specific domain. Finding the particular narratives is bottlenecked by knowing enough about the particular topic. Moreover, since it takes time to gather enough information about the different topics it makes it difficult to scale up to larger numbers of taxons.

Future work can explore automatic generation of the narratives given a list of tweets, and condense this list iteratively, and patch templates e.g., using pre-trained language models.

Directly modelling the initial argumentative text Finally, the approach we develop can operate on text that is claims or argument discourse units, but has no way of distinguishing between these or nonarguments. This precludes the model from being able to only predict a narrative if the text is indeed from the fine grained debate and can be tricked into providing narratives which the text doesn’t follow.

References

Binxuan Huang and Kathleen M. Carley. 2019. A large-scale empirical study of geotagging behavior on Twitter.


Haokun Liu, Derek Tam, Mohammed Muqeeth, Jay Mohta, Tenghao Huang, Mohit Bansal, and Colin Raffel. 2022. Few-shot parameter-efficient fine-tuning is better and cheaper than in-context learning.

Valentin Liévin, Christoffer Egeberg Hother, and Ole Winther. 2023. Can large language models reason about medical questions?


A Implementation Details

Here we describe the implementation details for fine-tuning the 3B T0 model for narrative prediction, in addition to using different ICL strategies for stance and aspect prediction. For all downstream tasks, we use the same AdamW optimiser with linear learning rate decay and weight decay. Finetuning details such as number of epochs and learning rate is reported in Table 10.

Aspect prediction For our aspect prediction models we use the standard BERT model to predict a sequence of BIO tokens. We tokenise a given sentence using the TreebankWordTokenizer from the nltk package available for the Python programming language. For the ICL setup we force the T0 model to consider only the words in the given sentence by tokenizing the sentence and feeding it into force_words_ids, additionally we also force the decoding step to not include stop words in addition to special characters like " that appear in the sentence.

During decoding, we set the temperature $\tau = 0.7$, top_p=0.9, number of beams equal to 5 to provide a variety of sentences following the same narrative.

Stance prediction For the Stance prediction we restrict decoding to one word only, and giving the model two choices for or against for the T0 model. For the baseline model we simply attach a LM head and do binary classification 0 = for and 1 = against.

Narrative prediction During finetuning we switch the standard T0 model out with T0-few with the LoRA setup and mainly keep the default hyperparameters but reduce the batch size to 4 and train a model for each topic for 10 epochs. Each model takes around 3hrs to train on the 10k training sentences. In addition to this setup we also include sentences that we generated sentences using the topic, predicted stance and aspect using the CTRL-UKP model, T0-3B, T5-flan-3B and 175B Bloom model. The tweets we predict the stance and aspect is from the test set. Using these attributes we can generate similar sentences to the test set to help enhance performance. We simply copy the target narratives as labels for the generated sentences and include them in the training dataset.

To give an example of the runtime for our code it takes 12 hours to complete 10 epochs for the T0-3B model using 1 TitanRTX-24GB and 1 Xeon E5-2620 v4 8c/16t - 2.1 GHz CPU, and 8 hrs using 1 A100-40GB and the same CPU for the T0-11B parameter model. We always have access to a minimum of 48GB of RAM but run our experiments using 64GB RAM.

B Search Query and Narrative Synonyms

<table>
<thead>
<tr>
<th>Topic</th>
<th>Search query</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGI</td>
<td>AGI replace humans, AI replace humans, AGI technology AGI threat, AI threat, AGI beyond human intelligence, AGI rule the world AI useful, AI better than people, society help AI, AI brain human brain better AI, AGI achieves human, AGI tech, human ethics AGI AI threat humanity, AI solves problem, AI help climate, AI hype bad AI hype up, AI hype good, AGI future good, AI as common sense AI trust bad, AI superhuman, AI wants things that are absurd to humans AI billionaire control, AI just tool, AI wealth concentration AI wealth inequality, automatic wealth inequality, AI if statements as unenforceable, AI make me laugh, AI art real AI meme, AI meme, AI problems live, AI lies our problems AI human nature, AI data now oil, AI coming for job</td>
</tr>
<tr>
<td>Alternative meat</td>
<td>stop subsidizing meat, alternative meat fake, alternative meat unhealthy meat is cruel, soy meat replacement, reduce meat consumption climate meat no sustainable, meat is unhealthy, food pyramid scheme substitue green nutrition, increase production of meat exempt meat production from carbon taxes, carbon tax to food production invest Meat alternatives, Meat alternatives subsidized Plant based food subsidized/introduce meatless mondays Vegetarian vegan food encouraging vegge diet, subsidize fruits vegetables meat overconsumption, Plant based food encourage, Meat alternatives encourage plant based food sustainable, plant food is great, fresh organic food is good meat alternative food is good, red meat is bad, animals are not ingredients can healthy food, raw food diet Joost meat alternative big pharma alternative meat, alternative meat forced plant based food processed, plant based food remove meat animals can meat humans too, cut plant save planet, eat meat saves plant, meat eating planet, alternative meat bugs</td>
</tr>
</tbody>
</table>

Table 11: Twitter search keywords

Table 11 - 13 lists the Twitter queries we used to retrieve the initial training data. Note that many of the words are either in their stem or shortened format in order to ensure a wider range of search results being returned. Per default twitter filter out sentences that does not contain tokens from the
For our crowdsourcing of narrative annotations and human evaluation we use Amazon Mechanical Turk. Workers had to take a qualification test, have an acceptance rate of at least 80% based on 5 question, be located within the US, have successfully completed more than 1000 HITs before and have an approval rate of 98%. We paid 1 dollar per HIT for the dataset task which is to classify one tweet into one in roughly 40 narrative categories. Initially time spend on a HIT is much higher than when they complete their 25th hit as workers learn to memories the categories. For the human evaluation we get consent from people whose data we are using though the Twitter Term of Services. The data collection procedure was approved by our internal ethics review board.

During our annotation of the narrative labels we discovered that the returned answers tend to be biased towards the top 10 first possible answers that could be selected in the HIT. To mitigate potential bias we manually went through the top 3 most frequent answer for each topic in the validation set and relabel the corresponding tweets.

Table 12: Twitter search keywords

Table 13: Twitter search keywords
D Crowdsourced Annotations

For this paper, gathering annotations has happened over three annotations rounds, each focusing on different sections of the paper.

D.1 Pretest

The first crowdsourcing task is that of a pretest, which is used to determine if workers are suitable for our main annotation mask. It is based on data from (Hansen and Hershcovich, 2022) and focuses on correctly classifying two different types of labels: Pro/con and evidence.

D.1.1 Pro/Con

Pro/con is a binary label. The tweet is annotated as (+1) for pro when a clear claim has a positive or supportive stance towards the topic. It is annotated as (-1) when it has a clearly antagonistic or attacking stance towards the topic. We exclude data for which has no clear stance.

Instructions for annotators: Given a tweet your task is to annotate it with its stance in relation to the topic. The stance is either pro or con (for or against a topic). In this case select pro if you find that the tweet is supportive towards the topic, and con if it is hostile instead. Remember that a tweet with hostile remarks can still be supportive of the topic, as we want to find the stance towards the topic and not the tweet itself.

D.1.2 Evidence

Evidence as a label has 6 classes. The tweet is annotated using any of the labels: Normative, Study, Expert, Fact, Anecdotal or unrelated/no evidence. The description for each of these labels are taken from (Hansen and Hershcovich, 2022): Anecdotal refers to "a description of an episode(s), centred on individual(s) or clearly located in place and/or in time." Expert refers to a "testimony by a person, group, committee, organisation with some known expertise / authority on the topic. Study refers to "results of a quantitative analysis of data, given as numbers, or as conclusions" Fact refers to "A known piece of information about the world without a clear source for the information" Normative refers to "an added description for a belief about the world" No evidence refers to "the tweet does contain evidence, but it is not related to the topic, or it does not have any evidence."

Instructions for annotators: The task is to annotate a tweet with the type of evidence it contains. Evidence is a statement used to support or attack a topic or claim. Evidence can be present in combination with a claim, or it can also be self-contained if it is just stating facts or referencing studies related to the topic. If the evidence is unrelated to the discussed topic, it is marked as unrelated. If you feel that multiple types of evidence is present in the tweet, choose the one that you think best describes the main piece of evidence in the tweet. Remember that your task is to annotate the type of evidence that is in the tweet regardless of your views and if the evidence is true or not.

D.2 Narrative annotation

The main crowdsourcing task of this paper is essentially claim classification. Given a tweet the workers determine if the tweet is a claim or argument with an evidence type that is not a study (as taken from the definition of study in (Hansen and Hershcovich, 2022)). Then if the tweet is a claim then they should select the most similar claim from a list of options. If no option is suitable, they should select "No claim in list is similar to the tweet".

Instructions for annotators: The task here is to annotate a tweet given a list of claims that the tweet might be similar to. Of course, each tweet can be relevant for more than one claim, but it can also be irrelevant and should be annotated as such. Therefore, given that the topic is {} select the claim which you find the tweet most similar to (regardless of your views on the list of claims, the topic and the tweet itself). Remember that the surrounding context of a tweet can be missing, and that people may be sarcastic.

D.3 Human evaluation of generated claims

The last crowdsourcing campaign is the human evaluation in which we evaluate how well a generated claim compared against the original claim (It is generated from the predicted stance and aspect from the original claim ). We follow primarily (Skitalinskaya et al., 2022) for definition of argument quality, meaning and fluency, but also (Schiller et al., 2021) for fluency and persuasiveness. These generated claims are then used for finetuning a LLM for improved narrative prediction.

Instructions for annotators: In this task, you will identify if a generated claim is similar to or has improved, without changing the overall meaning of the text. Each field contains a pair of tweets, one being the original and the other a synthetic tweet
that is trying to mimic it. Please rate each candidate along the following four perspectives: argument quality, fluency, meaning and persuasiveness.

Argument Quality has a scale from 1 to 5: 1 (notably worse than original), 2 (slightly worse), 3 (same as original), 4 (slightly improved), 5 (notably improved) Does the generated claim improve over the original claim? Things to look for include: specifying a fact, simplifying the sentence, adding clarity, adding additional information such as facts, adding, editing or removing links for external resources.

Meaning has a scale from 1 to 5: 1 (entirely different), 2 (substantial differences), 3 (moderate differences), 4 (minor differences), 5 (identical) Here we wish to measure if the generated claim have the same overall meaning as the original. Adding extra information that does change the objects or events described in the claim should not penalise the score.

Persuasiveness runs from 1 to 3. 1 (generated text less persuasive than original), 2 (equally persuasive), 3 (generated text is more persuasive) (choose one argument as being more persuasive or both as being equally persuasive.) Here we wish to measure if the generated claim is more useful in a debate about a certain topic than the original claim. Adding additional text that explains an event or fact more in depth should be rewarded.

Fluency runs from a scale form 1 to 3: 1 (major errors, disfluent), 2 (minor errors), 3 (fluent) Here we want you to to compare the generated sentence with the original one and ask if the sentence is written in fluent English and makes sense? You should consider rewarding the generated claim in case of improved grammar, spelling and punctuation of generated claim over the original claim.

E Narratives per topic

<table>
<thead>
<tr>
<th>Topic</th>
<th>Narrative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abortion</td>
<td>Abortion reduces crime</td>
</tr>
<tr>
<td></td>
<td>Abortion should not be allowed</td>
</tr>
<tr>
<td></td>
<td>Everyone has a right to life</td>
</tr>
<tr>
<td></td>
<td>A fetus is a real person</td>
</tr>
<tr>
<td></td>
<td>Abortion is painful for the fetus</td>
</tr>
<tr>
<td></td>
<td>Abortion is not murder</td>
</tr>
<tr>
<td></td>
<td>Abortion reduces the value of human life</td>
</tr>
<tr>
<td></td>
<td>Women that go through abortion face social stigma or guilt</td>
</tr>
<tr>
<td></td>
<td>Supporting abortion is societal pressure</td>
</tr>
<tr>
<td></td>
<td>Abortion gives mothers the option of giving birth to healthy children</td>
</tr>
<tr>
<td></td>
<td>No abortion option for poor women is injustice</td>
</tr>
<tr>
<td></td>
<td>A fetus is not a real person</td>
</tr>
<tr>
<td></td>
<td>Abortion is murder</td>
</tr>
<tr>
<td></td>
<td>Planned children lead better lives</td>
</tr>
<tr>
<td></td>
<td>Modern medicine makes abortion is less of a risk</td>
</tr>
<tr>
<td></td>
<td>Couples that cant get kids want to adopt</td>
</tr>
<tr>
<td></td>
<td>Do not have kids if you fear they will be born with defects</td>
</tr>
<tr>
<td></td>
<td>Fathers have no say if the mother wants abortion</td>
</tr>
<tr>
<td></td>
<td>Abortion is not painful for the fetus</td>
</tr>
<tr>
<td></td>
<td>Removing abortion can put some pregnant woman at risk</td>
</tr>
<tr>
<td></td>
<td>Women that abort have no dignity</td>
</tr>
<tr>
<td></td>
<td>Abortion leads to mental diseases</td>
</tr>
<tr>
<td></td>
<td>Abortion encourages more sex</td>
</tr>
<tr>
<td></td>
<td>Authority are against performing abortion</td>
</tr>
<tr>
<td></td>
<td>anti-abortion is counterproductive</td>
</tr>
<tr>
<td></td>
<td>Abortion is inhuman</td>
</tr>
<tr>
<td></td>
<td>restricting abortion enforce traditional gender stereotypes</td>
</tr>
<tr>
<td></td>
<td>women that have been raped should have right to abort</td>
</tr>
<tr>
<td></td>
<td>Abort is morally wrong</td>
</tr>
<tr>
<td></td>
<td>abortionists are in it for the money</td>
</tr>
<tr>
<td></td>
<td>Fetuses should be protected</td>
</tr>
<tr>
<td></td>
<td>Children who almost got aborted might feel rejected</td>
</tr>
<tr>
<td></td>
<td>pro-life views makes no sense</td>
</tr>
<tr>
<td></td>
<td>Parents must know if their child has an abortion</td>
</tr>
<tr>
<td>No claim in the list is describing the tweet</td>
<td></td>
</tr>
</tbody>
</table>

AGI

<table>
<thead>
<tr>
<th>Narrative</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI is just hype</td>
</tr>
<tr>
<td>AI art unlike human art does not have any value</td>
</tr>
<tr>
<td>AI is just if else statements</td>
</tr>
<tr>
<td>AI has no common sense</td>
</tr>
<tr>
<td>Current AI is not superhuman</td>
</tr>
<tr>
<td>AIs do not have empathy</td>
</tr>
<tr>
<td>AI is bad because it is not as good as human</td>
</tr>
<tr>
<td>AI can make you laugh</td>
</tr>
<tr>
<td>AI will not replace humans but augment them</td>
</tr>
<tr>
<td>AI is bad as it replaces artist</td>
</tr>
<tr>
<td>AGI is just a myth</td>
</tr>
<tr>
<td>AI is for the most part uncontrollable</td>
</tr>
<tr>
<td>AI will create more problems than it solves</td>
</tr>
<tr>
<td>You cannot trust AI</td>
</tr>
<tr>
<td>AI will not take your job</td>
</tr>
<tr>
<td>Data is important to make good AI</td>
</tr>
<tr>
<td>AGI will rule the world</td>
</tr>
<tr>
<td>We will get AGI sooner than expected</td>
</tr>
<tr>
<td>AI is a threat to humans</td>
</tr>
<tr>
<td>AI will fix our problems</td>
</tr>
<tr>
<td>AI cannot recreate human nuances</td>
</tr>
<tr>
<td>AI will take your job</td>
</tr>
<tr>
<td>AI will demotivate you from working</td>
</tr>
<tr>
<td>AI is stealing from artist</td>
</tr>
<tr>
<td>You cannot trust people who hype up AI</td>
</tr>
<tr>
<td>AI will help us solve climate change</td>
</tr>
<tr>
<td>AI will live up to its hype</td>
</tr>
<tr>
<td>AI is already superhuman</td>
</tr>
<tr>
<td>AI is just a tool</td>
</tr>
<tr>
<td>AI is power hungry just like the billionaires who control it</td>
</tr>
<tr>
<td>AI furthering the wealth inequality</td>
</tr>
<tr>
<td>AI is a general purpose technology like electricity</td>
</tr>
<tr>
<td>No claim in the list is describing the tweet</td>
</tr>
</tbody>
</table>

Table 14: First list of narratives
## Table 15: Second list of narratives

<table>
<thead>
<tr>
<th>Topic</th>
<th>Narrative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corporate culture</td>
<td>People see companies and big profit companies are equally unsatisfactory. Barriers to trade often lead to monopoly companies. A single company can control an industry. Companies simply need demos that do not ask questions. It is a clay court for motorcycles. Working minutes does not mean that company culture is not important. If you want great company culture you should hire good people with a good work ethic. Corporate culture is bad. Corporate culture is resistive. Companies can do what they want to do. Nowhere do you not see work. The side effects of business culture in each of the respective sampling of systems are getting lines on walls while you are getting food. People can get the work ethic and a bad work ethic. Victims of abuse are ignored and silenced. Companies that need roadside help that nobody wants to work in the public or get social profit must. NGOs are consumer in expression takeover. One company culture is great because they have free food. Corporate culture is not learned hard work. Art is consuming hard on our culture. Companies are anti-stress. Corporate culture is culture in the city by permanent. Getting a stable job is getting harder over time. A firm company culture does not care about your work life balance. Some other cultures are like a cult and are not healthy for you. Legitimately in the company they do everything. It is on the outside of the trouble. Parks from a company are endless. Young people prefer to suffer or stay in the workforce. Companies that make you feel like a family is the most important. Companies that say that you are a family is hausenishing you to comply. Good work gets everywhere. Corporate culture promoting genuine pay and perks in order to make workers dependent and exploitation legally in the company means absolutely nothing today and age. You can make a successful work life part of the team without being in the same physical space. An employee is an representative of a company and should respect the book and ask professionals work. There is no clear way to do it. Do not trust companies. Are there in the list is describing the resort.</td>
</tr>
<tr>
<td>Crypto</td>
<td>People will sell you but you are not rich crypto. Crypto is used mostly for pump and dump schemes. Crypto is used solely for pump and dump schemes. Crypto allows you to be happy, crypto avoids this problem. Crypto provides upwards mobility.</td>
</tr>
</tbody>
</table>

## Table 16: Third list of narratives

<table>
<thead>
<tr>
<th>Topic</th>
<th>Narrative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corporate culture</td>
<td>People see companies and big profit companies are equally unsatisfactory. Barriers to trade often lead to monopoly companies. A single company can control an industry. Companies simply need demos that do not ask questions. It is a clay court for motorcycles. Working minutes does not mean that company culture is not important. If you want great company culture you should hire good people with a good work ethic. Corporate culture is bad. Corporate culture is resistive. Companies can do what they want to do. Nowhere do you not see work. The side effects of business culture in each of the respective sampling of systems are getting lines on walls while you are getting food. People can get the work ethic and a bad work ethic. Victims of abuse are ignored and silenced. Companies that need roadside help that nobody wants to work in the public or get social profit must. NGOs are consumer in expression takeover. One company culture is great because they have free food. Corporate culture is not learned hard work. Art is consuming hard on our culture. Companies are anti-stress. Corporate culture is culture in the city by permanent. Getting a stable job is getting harder over time. A firm company culture does not care about your work life balance. Some other cultures are like a cult and are not healthy for you. Legitimately in the company they do everything. It is on the outside of the trouble. Parks from a company are endless. Young people prefer to suffer or stay in the workforce. Companies that make you feel like a family is the most important. Companies that say that you are a family is hausenishing you to comply. Good work gets everywhere. Corporate culture promoting genuine pay and perks in order to make workers dependent and exploitation legally in the company means absolutely nothing today and age. You can make a successful work life part of the team without being in the same physical space. An employee is an representative of a company and should respect the book and ask professionals work. There is no clear way to do it. Do not trust companies. Are there in the list is describing the resort.</td>
</tr>
</tbody>
</table>

| Cloning | Clones can perfect can give humans preferable qualities. Corporate culture promoting genuine pay and perks in order to make workers dependent and exploitation legally in the company means absolutely nothing today and age. You can make a successful work life part of the team without being in the same physical space. An employee is an representative of a company and should respect the book and ask professionals work. There is no clear way to do it. Do not trust companies. Are there in the list is describing the resort. |
breastfeeding is natural and is not politically correct
women are pressured into breastfeeding and gets stressed
breastfeeding is costly
breastfeeding is killing babies
It does not have to be all or nothing
the political right3 vote against bills to make things more expensive
breastfeeding and baby formula is risky
breastfeeding ends up in foreign countries instead of the US where it should be
Some people are allergic to breast milk and need formula
People hate babies when they make abstinence illegal and remove baby formula
baby food industry is promoting propaganda
breastfeeding can cause IRD
baby formula is poisonous
baby formula is good as fathers can feed their baby
baby formula is hot if you cannot breastfeed
It is unwise to impose breast formula in an economic crisis
breast milk is best
The political left try to remove their baby
People are recalling baby formula in other countries for higher prices
Do what is best for you and feed your baby
The baby formula shortage is one big scam
Some people cannot tolerate formula and need milk
breast milk has a lot of antibodies that can help the baby fight off infection
people publicly shame women who breastfeed in public
mental health is more important than breast feeding
breastfeeding is healthier than baby formula
baby formulas can cause infections
breastfeeding will lower risk of breast cancer
The political right hate a cause a baby formula shortage
breast in best campaign causes anxiety in moms who cannot breastfeed
No claim in the list is describing the tweet

Influencers
Influencers can be damaged by everything they say
social media in the west is like opium for kids
influencers just want to get rich quick
influence marketing are not authentic
social media content is not sustainable
influencers are creative
influencers care too much money
social media people are toxic and rude
an influencer is a social media celebrity
Formal jobs are boring
Influencing is indeed hard work
influencers understand the use cases of products and want to help
you should not quit your job and become an influencer
social media people are just plagiarizing other people
Becoming an influencer allows you to live the good life
influencers do not know hard work
influencers want to be their own boss
dealing with hate is part of a social media job
the numbers of followers do not make you successful
influencers are not respected
influencers are not an adult job
jealous youth are spending too much time on social media platforms
If you have less followers you should get a job
influencers are wasting their time
being an influencer is easy
influencer is not a real job title
No claim in the list is describing the tweet

Mental Health in sports
male-dominated sports are toxic for women
In sports people get into habits when mental health declines
sports help you alleviate stress
The money made in sports go to mental health organizations or admin staff
teams sports is a brutal business
female athletes are not protected
When athletes get it real they the media
sports in like religion is hard for your mental health
In sports racism and mental health issues goes hand in hand
athletes do not have any problems
athletes are not as bad as bad they think
athletes do not have real mental health problems
Work long hours, don’t take days off
Be tough, vulnerability in weaknesses
Emotional stress is an early sign to start getting help is stigmatizing
elite athletes have similar genetic advantages
sports athletes are manipulated
mental health is not masculine
athletes are only thriving professionally if they thrive personally
mental health should not be treated like the flu
athletes should not participate in male or female sports
be ashamed if talking about mental health
sports athletes are depressed
you do as your told as an athlete
athletes cause mental health issues in sports
athletes should not worry because they have a lot of money
talking about mental health is showing weakness
athletes that speak up about mental health issues are silenced
No claim in the list is describing the tweet

Nuclear Energy
Nuclear reactor is easy to control
Deciding what to do with regards to long-term disposal of nuclear energy waste is difficult
Nuclear energy will be available for use longer than oil or coal
Nuclear energy will continue to be the environment
Nuclear energy is dangerous
Nuclear energy can give an unlimited energy
Nuclear energy waste can be recycled
Nuclear energy is good
Using nuclear energy to solve problems that arise is logical
Nuclear energy leads to more violence
nuclear power produces carbon-free energy
nuclear energy is not efficient
nuclear power is financially burdensome
There is no significant risk with nuclear energy that cannot be said about other agents as well
nuclear energy is dirty
nuclear energy makes poor nations dependent on rich nations
Every country can use nuclear energy unlike everything else
nuclear energy relies too heavily on subsidies
There is not a good plan for storing or disposing of nuclear energy waste so we should use it
Nuclear plants only produce electricity and cannot replace oil and gas
Using nuclear power leads nuclear war
nuclear energy is a more viable option than nuclear energy
Nuclear energy is favored by certain social structures like capitalism
decentralized nuclear energy production is efficient
Nuclear power is needed to stabilize climate change
Nuclear energy should not even be considered as an energy source
Nuclear energy is much more harmful than beneficial
Green energy will make nuclear energy obsolete
Nuclear energy will increase the cancer and illnesses
nuclear energy is not renewable energy
Nuclear reactors are vulnerable to terrorist attack
nuclear energy is more reliable than renewable energy sources like solar
No claim in the list is describing the tweet

Transport
trains are better than flights
public transportation is more sustainable for rural areas
public transportation in nuclear car is better
cars give you the freedom of independence
buses are safer than cars
fastest drivers equal safer streets
public transportation is comfortable
transportation in public transit is less pollutions
trains are better for the climate
public transportation has no personal space
It is important that public transit works
cars are good when there is no alternative
planes are better than trains
public transportation is for poor people
public transportation is hidden from disease
highways are too predictable
public transit only works if affordable
using bikes are dangerous
public transportation is filled with germs
cars are worse than buses as it carries less people
buses are better than care
public transportation is a good business
using car and road that the poor people
cars are for rich people
public transportation is nice in cities but car-centric infrastructure is bad
trains are too expensive
No good transportation is in a reflection of the government
climbing will decrease car traffic
public transit is not profitable
public transportation is good
No claim in the list is describing the tweet