# Fact-checking Epistemology Annotation Guidelines

Michael Schlichtkrull, Nedjma Ousidhoum, Andreas Vlachos
Department of Computer Science and Technology
University of Cambridge
mss84,ndo24,av308@cam.ac.uk

March 31, 2023

To study the epistemology of automated fact-checking, we annotate the narratives of 100 highly cited research papers.

We define a two-step annotation scheme: 1) a paragraph-level annotation and 2) a discourse-level narrative annotation. In the paragraph-level annotation, we extract quotes related to the goal and the methodology presented by identifying identify the a) data subjects, b) actors, c) model means, d) application means, and e) epistemic ends. Then, based on the identified elements, we extract the implied narratives in the discourse-level annotation.

We went through three-round annotation process. First, we annotated a small pilot set of five papers to define an annotation scheme. Then, working inductively, we annotated the full set of (100) research papers. That is, if, during the annotation, we encounter a means, end, actor, or narrative that does not fit any of our given categories, we introduce a new one. We then discussed our annotations and unified our set of introduced categories into those discussed in Sections 2 and 3. We further created a flowchart to help us move from paragraph-level to discourse-level annotation in a structured way (see Flowchart 5.3). Finally, we re-annotated all 100 papers based on our unified set of categories. The final set of criteria can be found in the sections below.

## 1 Paper Selection

We collect research papers on automated fact-checking and related tasks (e.g., rumor detection). Working from the GitHub repository of fact-checking papers published alongside Guo et al. 2022 [2], we select the 100 most cited. We examine the introductions and abstracts of each paper to extract quotes

# 2 Paragraph-Level Annotation

## 2.1 Quotes

We extract the quotes/paragraphs from the introduction, yet, if a piece of information is missing, we further look at the abstract. We only examine epistemic quotes, i.e., related to knowledge. We sort them into quotes answering questions about narratives, i.e., the *why* and *what* of the paper.

## 2.2 Citation Support (Origins)

We extract any backing used for the narrative. This could be a citation for why the task/paper is important. E.g.,

- a scientific article,
- a newspaper article/poll,
- a previous NLP paper,
- an anecdote,
- sense of threat (nothing).

## 2.3 Data subjects

Based on *"who did what to whom for whom"*, we extract the subjects/people whose texts are fact-checked.They can be:

- journalists,
- citizen journalists,
- social media users,
- technical writers,
- public figures/politicians,
- product reviewers.

**Social media users** Covers any (potentially anonymous) contributors to social media as long as they are not explicitly public figures. This category includes people who comment on forum, Twitter/Facebook users, or editors of Wikipedia and other collaborative writing projects.

**Journalists** Covers professional journalists, including fact-checkers, and anyone writing at a publishing house. This category also includes online publishers, but not collectives of amateur sleuths, e.g., Bellingcat. Organizations pretending to be journalists, e.g., satire sites and fake news websites, are also counted within this category.

**Politicians & Public Figures** Covers any public figure, such as a politician or an actor. Analysed data could be media releases, interviews, speeches, or similar.

**Citizen Journalists** Covers amateur journalists who take on the same work as professionals without funding or traditional education, e.g., bloggers, social media users, and collectives such as Bellingcat.

**Product Reviewers** Covers specifically product reviewers on sites such as Amazon, Trustpilot, where the purpose of the commenter is to describe and rate a product.

**Technical Writers** A few papers suggest scientists and other writers of technical documents should use fact-checking to, e.g., spot errors in their articles before publication. Along with scientific writers, this also covers writers of technical documents for areas such as business or health, as well as lawyers and clerks seeking to ensure consistency within legal documents.

## 2.4 Data actors

The people who are supposed to **act** on the model outputs, such as journalists, social media moderators, or media users. From our data, they can be :

- professional journalists,
- citizen journalists,
- social media moderators,
- scientists,
- media consumers,
- technical writers,
- engineers and curators.
- law enforcement,
- algorithm.

**Professional Journalists** Covers professional journalists, including fact-checkers, and anything written at a publishing house. This category does not include amateur sleuths, e.g., Bellingcat, or bloggers doing the work of journalists.

**Citizen Journalists** Covers amateur journalists who take on the same work as professionals without funding or traditional education, e.g., bloggers, social media users, and collectives such as Bellingcat.

**Social Media Moderators** Covers people hired to moderate social media spaces. Applicable only when it is explicitly stated that a human employee should act on the model outputs.

**Scientists** Covers scientists, as well as any other actors who would use model outputs for scientific research. E.g., to analyse data with the express purpose of learning something about it, not acting on the model decisions.

**Media Consumers** Covers ordinary people who consume the content to which the fact-checking system is supposed to be applied. Only applicable when the consumer is directly understood (possibly implicitly) to use the system, e.g., in the form of a browser extension. Does not cover cases where model decisions are shown to social media users through, e.g., warning labels, the decision to use the tool must be in the hands of the consumer.

**Engineers and Curators** Covers cases, where engineers or curators are maintaining a knowledge base of some kind, (e.g., Wikipedia), are intended to use the model outputs in their work.

**Law Enforcement** Covers cases where law enforcement agents, e.g., police officers, intelligence agents, or judges, are intended to act on the model outputs.

**Algorithm** Covers the cases of fully automated systems that act on the model outputs, e.g., remove posts based on the model's predictions.

## 2.5 Model owners

The entities who **own** the models, or institutions who employ those who **act** on the model outputs. For instance, fully automated moderation systems are owned by the companies expected to use the models, e.g., social media companies. From our data, they can be :

- media companies,
- social media companies,
- law enforcement.

**Media companies** Covers professional (non-amateur) media companies employing journalists and fact-checkers.

**Social Media Companies** Covers social media companies more generally, including the engineers working to maintain the social network. Applicable when decisions will be made automatically based on the model's decisions, or when it is unclear.

**Law Enforcement** Covers cases where law enforcement agents, e.g., police officers, intelligence agents, or judges, are intended to act on the model outputs.

## 2.6 Modeling (ML) Means

What concretely do the authors propose to do in terms of machine learning models? E.g., classifying claims, finding evidence, or similar. From our data, these can be:

- Classify/score veracity,
- classify/score stance,
- evidence retrieval,
- justification/explanation production,
- corpora analysis,
- data collection,
- human in the loop,
- generate claims.

**Classify/score Veracity** When the authors propose to classify the *veracity* of claims, i.e., whether or not the claim is true (or supported by evidence).

**Classify/score Stance** When the authors propose to classify the *stance* of evidence, i.e., whether or not an evidence document takes a positive or negative view on a particular subject or claim.

**Evidence Retrieval** When the authors propose evidence retrieval as a mean to reach their goal.

**Provide Justifications** When the authors propose generating explanations/justifications to reach their goal.

**Human in the loop** When there are human actors involved in the solution/main process described in the paper.

**Corpora Analysis** When the authors propose using data analytics or any sort of corpora analysis in their methodology.

**Data Collection** When the authors propose collecting data to solve a given problem.

**Generate claims** When the authors propose generating claims, e.g., to produce additional misinformation to train on.

## 2.7 Application Means

How do the authors want to use what is developed in the paper to accomplish a specific goal (e.g., reduce the spread of misinformation)? For example, this could be by deploying automated systems to show social media users content warnings about claims which might be false (along with, potentially, evidence for their falsity). From our data, we identify the following suggestions:

- identify claims,
- triage claims,
- supplant human fact-checkers,
- gather and present evidence,
- identify multimodal inconsistencies,
- automated removal,
- provide labels/veracity scores,
- provide aggregates of social media comments,
- filter system outputs,
- maintain consistency with KB,
- analyse data,
- produce misinformation,
- vague persuasion.

**Identify claims** There are many claims on the internet and most of them are not misinformative. ML should be deployed to find the misinformative ones.

**Triage claims** Fact-checkers, content moderators, and similar have many claims to deal with. ML should be deployed to rank them, so more costly actors focus on the most important ones first.

**Supplant Human fact-checkers** Replace human fact-checkers entirely, or at least partially by automatically handling some claims. If the intention is a human-in-the-loop system, *supplanting human fact-checkers* is not the means.

**Gather and present evidence** An ML model should find relevant evidence supporting/refuting a claim, and show it to a human. It can also involve generating justifications for how the evidence relates to the claim.

**Identify multimodal inconsistencies** For multimodal misinformation, an important tool identifies mismatches between modalities. ML models should do this, then show the results to humans.

**Automated removal** ML models should automatically remove claims from e.g., social media platforms, with no human involvement.

**Provide labels/veracity scores** ML models should provide indications of truth value to media consumers, either in the form of labels or veracity scores. The consumers can, but do not have to be, the data actors. If a company decides to *always* show a score, the company is the data actor.

**Present aggregates of social media comments** ML models should aggregate the stance/evidence presented in the comments to a potentially misinformative claim, as a way of summarizing points made by humans in favor of or against the claim.

**Filter system outputs** Other NLP/ML models, e.g. LLMs, struggle to produce truthful outputs. Fact-checking models should filter or re-rank their output. This also includes extractive models, e.g., relation extraction systems, that are coupled with a fact-checker to only return (or add to a KB) things the fact-checker accepts as truthful.

**Maintain consistency with KB** Internal knowledge bases (defined loosely, including textual ones, such as Wikipedia) can be used as a source of truth to prevent the data actor from publishing untruths. This could be at writing time, e.g., as a writing assistant, or it could be by continuously keeping an article published online up to date.

**Analyse data** Fact-checking models and datasets should be used for research purposes, e.g., to analyse misinformative text to get a better understanding of how misinformation spreads.

**Produce misinformation** A machine learning model that produces misinformation can be used *against* misinformation, for example, to generate adversarial data, or to show people what such models are capable of to innoculate them against it when they encounter it in the wild.

**Vague persuasion** ML models should somehow convince people to change their opinions on claims, e.g., by providing warning labels or presenting evidence. The mechanism is *not specified*, though.

## 2.8 Ends

The purpose or ultimate aim of the approach. What do the authors want to accomplish? From our data:

- limit misinformation,
- limit AI-generated misinformation,
- increase the veracity of published content,
- develop knowledge of NLP/language,
- avoid biases of human fact-checkers.
- detect falsehood for law enforcement

**Limit misinformation** The ultimate aim of the paper is to prevent misinformation from spreading or to limit its influence.

**Limit AI-generated misinformation** The ultimate aim of the paper is to prevent AI-generated misinformation from spreading or to limit its influence.

**Increase veracity of published content** Some systems are intended to be used by publishers and writers before their content is made public – or applied to continuously keep published content up to date. The aim here is not to limit already spreading misinformation but to keep people from accidentally publishing untrue things. Published content can be small data, like a single article, a large collection of data, like Wikipedia, or a knowledge base.

**Develop knowledge of NLP/language** Automated fact-checking is a difficult problem. By studying it, we can learn more about how language works, and how to build models that interpret semantics.

**Avoid biases of human fact-checkers** I.e., develop "super-human" fact-checking.

**Detect falsehood for law enforcement** One proposed use case for fact-checking and similar technologies is as truth-telling systems for law enforcement, e.g. in courtrooms.

## 2.9 Important Note

In this annotation stage we do not extract implied statements, we rely on extracted quotes and do not interpret them.

# 3 Narratives

At the discourse level, we extract the epistemic narratives present in the paper. We envision narratives as discourse structures combining the (modeling/application) means, ends, data actors, and data subjects extracted at the paragraph level. Note that we use fictional examples below to avoid biasing annotators' decisions on specific papers.

The narratives can be:

- vague identification,
- vague debunking,
- vague opposition,
- assisted content moderation,
- automatic content moderation,
- assisted media consumption,
- assisted internal fact-checking,
- automated external fact-checking,
- assisted knowledge curation,
- scientific curiosity.
- truth-telling for law enforcement
- vague moderation,
- adversarial research.

We do not account for implied paragraph-level narratives but only for annotated elements that *are* in the quotes. As we extract multiple quotes, papers may have more than one narrative present.

**Vague Identification** The paper mentions identifying or detecting misinformative claims as the means, and limiting misinformation as the end. However, it is not clear how the authors intend to accomplish that end using those means. Typically, there are no data actors – it is also not clear who should act on the model's predictions.

*Vague identification* Applies only in cases where the authors say that they want to identify or detect misinformation without saying what they are going to do with that afterward, e.g., in rumor detection papers where they claim that they want to detect rumors but we do not know what they want to do with this identification or classification labels.

E.g., *"Misinformation will be fought by automatically identifying rumors."*

**Vague Debunking** When the authors propose to assist the external fact-checking process without explaining how.

It is clear that the mechanism is supposed to follow what fact-checkers are currently doing, but not where or how the ML model will be used in this process. Furthermore, it is *unclear* whether the entire process will be automated. For example, the paper may suggest that an automated fact-checking model should be used by fact-checkers, but it is not clear *how* the model assists in that. A common warning sign is a mismatch between ML means and application means: classification is not really useful for fact-checkers (or consumers of fact-checks), and so papers suggesting to use those either to assist or automate fact-checking are often not clear in how.

E.g., *"Models will be used for debunking (but not how)".*

**Vague Opposition** Restricted to cases without any application means (model means and no application means in contrast to vague identification and vague debunking). E.g., *a machine learning/automated system will reduce the spread of misinformation.*

When the paper presents a narrative of vague opposition to misinformation. The *end* is to limit the spread or influence of misinformation, and the *ML means* are, for example, to classify claims. However, the connection between means and ends is left unmentioned, and epistemic actors are typically absent. An impression is given that the development of automated fact-checking will limit the spread of misinformation, but the link between the two is left unstated. A (fictional) example follows below:

*"Misinformation is a major societal problem, eroding community trust and costing lives by e.g., inducing hesitance to adopt life-saving vaccines. It is therefore of paramount importance that the spread of false information is stopped. Automated fact-checking – that is, the automatic classification of claim veracity – represents one solution to this critical problem."*

**Assisted Content Moderation** If the paper proposes the deployment of automated fact-checking as a tool to assist content moderators on social media platforms. Here, the *end* is to limit the spread of misinformation on social media platforms, the *means* is to provide suggestions for posts to delete (along with, potentially, evidence for why they might be false), and the *actors* are human moderators who make the final choice on whether posts should be deleted or not. A (fictional) example follows below:

*"Social media is rife with misinformation, eroding community trust and costing lives by e.g., inducing hesitance to adopt life-saving vaccines. One solution is for moderators to remove information deemed false. However, with the number of posts made every day on social networks, this strategy is too costly. In this paper, we develop an automated system for filtering claims, helping moderators quickly discover and make decisions on circulating claims."*

**Automatic Content Moderation (replaces human content moderators)** If the paper analysed proposes a similar content moderation strategy, but instead of assisting human moderators, it suggests replacing them entirely. In this case, the *end* is to limit the spread of misinformation on social media platforms, the *means* is to deploy classifiers to truth-tell claims and remove any labeled false, and the *actors* are the executives and engineers at social media companies who deploy and make decisions about such systems. A (fictional) example follows below:

*"Social media is rife with misinformation, eroding community trust and costing lives by e.g. inducing hesitance to adopt life-saving vaccines. One solution is for moderators to remove information deemed false. However, with the number of posts made every day on social networks, this strategy is too costly. In this paper, we develop an automated system for detecting false claims, which can serve as a first line of defense against misinformation."*

**Assisted Media Consumption** If the paper proposes to deploy automated fact-checking as an assistive tool for *consumers* of information, either as a layer adding extra information to social media posts or as a standalone site where claims can be tested. In this case, the *end* is either to limit the influence of misinformation or to induce veracious mental states in users; the *means* is to deploy automated systems to warn about claims which might be false (along with, potentially, evidence for why their falsity); and the *actors* are social media users or information seekers in general. The actors could also be social media companies who show information produced by fact-checking assistants to users as an integral part of their UI. A (fictional) example follows below:

*"Misinformation is a major societal problem, eroding community trust and costing lives by e.g. inducing hesitance to adopt life-saving vaccines. It is therefore of paramount importance that the spread of false information is stopped. Studies have shown that many people adopt beliefs without doing due diligence on the information they receive. Automated fact-checking – that is, the automatic classification of claim veracity – could via e.g., a plugin be deployed to warn social media users about potentially false claims."*

**Assisted Internal Fact-checking** When the paper proposes to deploy automated fact-checking as an assistive tool for journalists, deployed internally. Here, the *end* is to increase the veracity of published information, the *means* is to deploy automated systems to warn about claims which might be false (along with, potentially, evidence for why their falsity), and the *actors* are journalists employed at traditional publishing houses. A (fictional) example follows below:

*"Research is a fundamental task in journalism, conducted to ensure published information is truthful and to protect the publisher from libel suits. This is a crucial step, which journalists – strained by the advent of the 24-hour news cycle – increasingly skip. Given a trusted source of evidence documents, such as LexisNexis, much of the grunt work of research could be handled by automated fact-checkers, leaving journalists free to tackle the hardest parts, e.g. double-*

*checking information with sources."*

**Assisted External Fact-checking** When the paper proposes to deploy automated fact-checking as an assistive tool for journalists, deployed for external fact-checking. Here, the *end* is to limit the influence of misinformation, the *means* is either to speed up the production of counter-messaging by surfacing relevant evidence or to direct journalists to the most problematic currently circulating claims, and the *actors* are journalists employed at fact-checking organizations such as Full Fact.

This is restricted to when the paper proposes to improve one or multiple components in the automated fact-checking pipeline rather than the whole pipeline/end-to-end system (in the latter case it becomes *automated external fact-checking*, see following paragraph). I.e., when it is human-in-the-loop, then it is assisted fact-checking but not automated.

A (fictional) example follows below:

*"Misinformation is a major societal problem, eroding community trust and costing lives by e.g., inducing hesitance to adopt life-saving vaccines. An important way to fight misinformation is the production of relevant counter-messaging, i.e., the work done by organizations such as Full Fact or PolitiFact. With the number of false claims published on social media every hour, it is not feasible for human journalists to debunk them all. Journalists could use automated fact-checking to triage incoming claims to limit the workload, or to quickly surface relevant evidence while producing articles."*

**Automated external fact-checking** Only when the authors say it explicitly, otherwise it is vague debunking (see previous paragraph).

I.e., when the paper proposes to *fully* automate the fact-checking process without a human-in-the-loop step.

**Assisted Citizen Journalism** When the paper proposes to deploy automated fact-checking as an assistive tool for regular people to produce counter-messaging against misinformation they encounter. That is, it would be used as a tool to enable *citizen journalists*. Here, the *end* is to limit the influence of misinformation, the *means* are to speed up the production of counter-messaging by surfacing relevant evidence, and the *actors* are ordinary citizens. A (fictional) example follows below:

*"Misinformation is a major societal problem, eroding community trust and costing lives by e.g., inducing hesitance to adopt life-saving vaccines. An important way to fight misinformation is the production of relevant counter-messaging, i.e., the work done by organizations such as Full Fact or PolitiFact. With the number of false claims published on social media every hour, it is not feasible for professional journalists to debunk them all. However, with the rise of citizen journalist collectives such as e.g., Bellingcat, groups of regular internet users can step in to supplement the professionals. Quickly finding relevant evidence to produce counter-messaging can be a difficult task, even for professional journalists;*

*however, automated fact-checking may make the task accessible to many people."*
influential

**Assisted Knowledge Curation** When the paper proposes fact-checking primarily as a component filtering the information kept in some curated knowledge vault, including graph-based knowledge bases as well as text-based collections such as Wikipedia. Here, the *end* is to increase the veracity of the knowledge vault, the *means* is to use automated fact-checking as an additional truth-telling layer that prevents disputed facts from being added (or interrogates already added facts), and the *actors* are typically the engineers who maintain the knowledge base. A (fictional) example follows below:

*"Knowledge bases fuel many real-world NLP applications, e.g., question answering. The maintenance of knowledge bases is an expensive process, yet as new facts appear in the world knowledge bases must be kept up-to-date. Automated triple extraction from e.g., news data has been proposed as an alternative to human annotators; yet, the quality remains low. Automated fact-checking systems, which verify facts against trusted knowledge sources, could be used to prevent highly disputed facts from being entered – or ensure that new facts are consistent with the existing knowledge."*

**Scientific Curiosity** When the authors of the paper justify their projects purely based on scientific curiosity. While differing strongly from the other narratives presented here, this is still a virtue epistemic narrative, concerned with the production of *good knowledge*. Here, the *end* is to increase scientific knowledge of semantics; the *means* is to learn how to build automated systems that mimic human fact-checkers, a process theorised to yield knowledge about the construction of meaning; and the *actors* are scientists in natural language processing and adjacent fields. A (fictional) example follows below:

*"Journalistic fact-checking is a difficult task, requiring reasoning about disputed claims that fool sufficiently many humans to warrant professional attention. For systems to mimic fact-checking to a substantial degree, significant semantic understanding is necessary. As such, automated fact-checking is an ideally suited field to develop and test new models for natural language understanding."*

# 4 Feasability Support

We annotate narratives with any support for their feasibility given in the paper. Vague narratives are by definition not supported – if the narrative is clear enough to design e.g., a user study to test if the means are an effective way to reach the end, it is not vague. We use the following categories:

**Scientific research** The feasibility of the narrative is supported by reference to scientific studies. This only applies if the entire narrative is supported, i.e., a scientific paper is supported for the means being a feasible way to reach the end.

This Does NOT apply if only e.g., the dangers of misinformation are supported – the proposed means being an effective strategy MUST be entirely supported.

For example, the crowdworkers' study in the FactCheckingBriefs (`https://aclanthology.org/2020.emnlp-main.580/`) could be cited to demonstrate that giving people evidence increases their accuracy on veracity judgments. This applies even if the cited paper does not support what the paper claims it supports, e.g., if someone cites FEVER for fully automated fact-checking being an effective strategy to reduce the spread of misinformation. This almost always applies to narratives of scientific curiosity, as researchers tend to show their research is a useful strategy for studying what they investigate by writing a related works section.

**News Media** The feasibility of the narrative is supported by a reference to a newspaper article. This occurs in the same cases as scientific research, except the citation is to a newspaper article rather than a scientific article.

**Automation** The feasibility of the narrative is supported by reference to the studied task currently working as a human task. It is assumed that full or partial automation will work, and will be effective – i.e., no reference to studies testing whether the proposed strategy helps humans in the loop is given.

**Example** *"While the number of organizations performing fact-checking is growing, these efforts cannot keep up with the pace at which false claims are being produced, including also clickbait (Karadzhov et al., 2017a), hoaxes (Rashkin et al., 2017), and satire (Hardalov et al., 2016). Hence, there is need for automatic fact checking."* [1]

**Vague Community** The feasibility of the narrative is supported by reference to the preference of practitioners, usually without citing a scientific paper. This includes phrases like *"many people think this is a good idea"*.

**Example** *"Many favor a two-step approach where fake news items are detected and then countermeasures are implemented to foreclose rumors and to discourage repetition."* [3]

**Sense of Threat** The feasibility of the narrative is not directly supported. However, the paper argues that the strategy represented by the narrative (e.g., automated content moderation) must be done, or bad things will happen. The bad thing is something other than human workers not being able to keep up with their workload, as the narrative otherwise falls under automation.

**Example** *"An abundance of incorrect information can plant wrong beliefs in individual citizens and lead to a misinformed public, undermining the democratic process. In this context, technology to automate fact-checking and source verification (Vlachos and Riedel, 2014) is of great interest to both media consumers and publishers."* [4]

# 5 Paper Metadata

## 5.1 The type of the paper

Survey, dataset, model description, task description, other (name).

## 5.2 The subfield/task

Can be fact-checking, rumor detection, misinformation, disinformation, or other (name the task).

## 5.3 The subtask

The part of the pipeline in which the model interacts, e.g., claim detection, evidence retrieval, verdict prediction, justification production.

# References


[1] R. Baly, M. Mohtarami, J. Glass, L. Màrquez, A. Moschitti, and P. Nakov. Integrating stance detection and fact checking in a unified corpus. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 21–27, New Orleans, Louisiana, June 2018. Association for Computational Linguistics.

[2] Z. Guo, M. Schlichtkrull, and A. Vlachos. A survey on automated fact-checking. *Transactions of the Association for Computational Linguistics*, 10:178–206, 2022.

[3] M. Potthast, J. Kiesel, K. Reinartz, J. Bevendorff, and B. Stein. A stylometric inquiry into hyperpartisan and fake news. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 231–240, Melbourne, Australia, July 2018. Association for Computational Linguistics.

[4] T. Yoneda, J. Mitchell, J. Welbl, P. Stenetorp, and S. Riedel. UCL machine reading group: Four factor framework for fact finding (HexaF). In *Proceedings of the First Workshop on Fact Extraction and VERification (FEVER)*, pages 97–102, Brussels, Belgium, Nov. 2018. Association for Computational Linguistics.

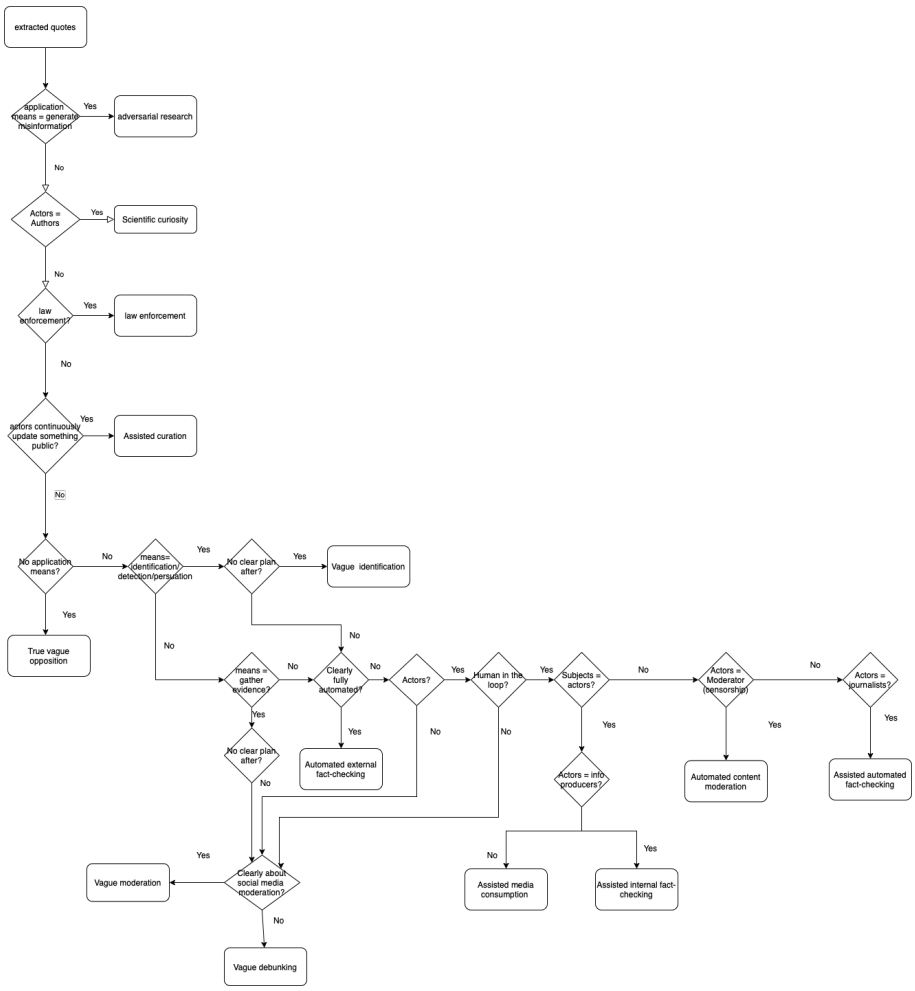


Figure 1: Flowchart of the Narrative Annotations.

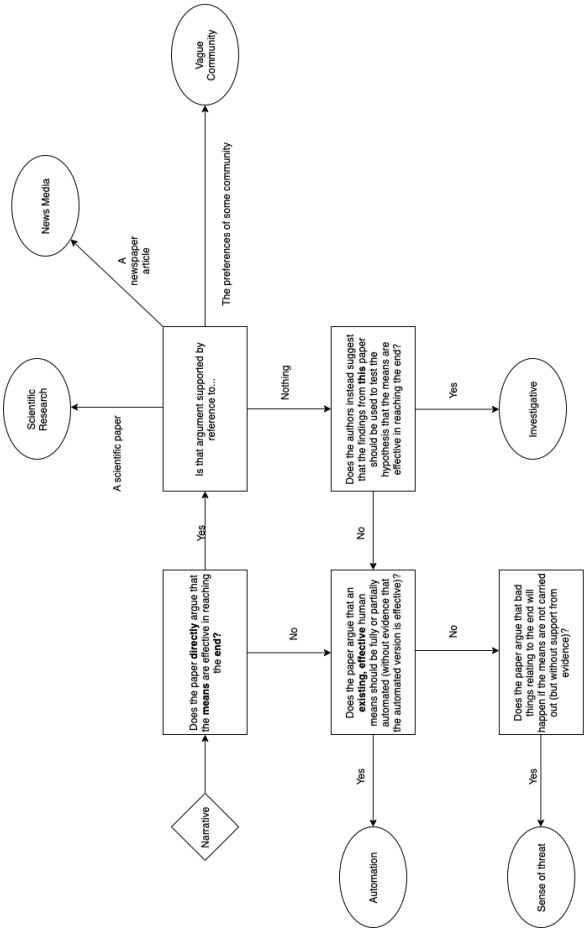


Figure 2: Flowchart of the Feasability Support Annotations.