

# Guideline for AVeriTeC annotators

v0.2 - February 25, 2022

## 1 Introduction

We aim to construct a dataset for automated fact-checking with the following guiding principles. First, we intend to decompose the evidence retrieval process into multiple steps, annotating each individual step as a question-answer pair (see Figure 1). Second, our dataset will be constructed from real-world claims previously checked by journalistic organisations, rather than the artificially created claims used in prior work.

Decomposing claim verification into generation and answering of questions allows us to break complex real-world claims down to their components, simplifying the task. For example, in Figure 1, verifying the claim requires knowing the salary of the health commissioner, the governor, the vice president and Dr. Fauci, so that they can be compared. Four separate questions about salary need to be asked in order to reach a verdict (i.e. that the claim is *supported*).

By decomposing the evidence retrieval process in this way, we also produce a natural way for systems to justify their verdicts and explain their reasoning to users. In addition to this, we annotate claims with a final justification, providing a textual explanation of how to combine the retrieved answers to reach a verdict. This allows users to follow each step of the retrieval and verification processes, and so understand the reasoning employed by the system.

**Claim:** Biden lead disappears in NV, AZ, GA, PA on 11 November 2020.

**Q1:** Which media project Biden will win in Nevada?  
**A1:** ABC News, CBS News, NBC News, CNN, Fox News, Decision Desk HQ, Associated Press, Reuters, and New York Times.

**Q2:** Which media project Biden will win in Arizona?  
**A2:** Fox News and Associated Pre.

**Q3:** Which media project Biden will win in Georgia?  
**A3:** None.

**Q4:** Which media project Biden will win in Pennsylvania?  
**A4:** ABC News, CBS News, NBC News, CNN, Fox News, Decision Desk HQ, Associated Press, Reuters, and New York Times.

**Verdict:** Refuted  
**Justification:** Many media organizations believe Biden will win in NV, AZ, and PA. As such, his lead has not disappeared.

Figure 1: Example claim and question answer pairs.

The annotation consists of the following three phases:

1. Claim Normalization.
2. Question Generation.
3. Quality Control.

Each claim should be annotated by different annotators in each phase. An annotator can participate *in* all three phases, but they will be assigned different claims.

**Warning!** Some components of the AVeriTeC may not render correctly in some browsers, specifically *Opera Mini*. If this is an issue we recommend trying another browser, e.g. Firefox, Chrome, Safari, or regular Opera.

## 2 Sign In

Each annotator will have received an **ID** and a **Password** with the access link to the annotation server. The password can be changed after logging into the interface.

### Important!

- Make sure to log out at the end of the session!
- Do not open multiple tabs/windows of the AVeriTeC annotation tool. Always use only one window during annotation!



Figure 2: Interface of the control panel. (1) Button for changing the password. (2) Button for logout. (3) Start the annotation for this phase. Here is Phase 1 Claim Normalization. (4) The left number shows how many claims have been annotated and the right number shows how many claims are assigned for the current annotator at this phase.

After clicking the **START NEXT** button, the annotation phase will start. If an annotator is new to the current phase, the interface will provide a guided tour as in Figure 3 for that phase. Please read the hints provided by the tour guide carefully before the annotation.

## 3 Phase 1: Claim Normalization

In the first phase, annotators collect metadata about the claims, and produce a normalized version of each claim, as shown in Figure 4. The first step is to identify the claim(s) in the fact-checking article. Often, this can be found either in the headline or explicitly some other place in the fact-checking article. In some cases, there may be a discrepancy between the article and the original claim (e.g. the original claim could be “*there are 30 days in March*”, while the fact-checking article might have the headline “*actually, there are 31 days in March*”). In those cases, it is important to use the *original* version of the claim.

If there is ambiguity in the article over the exact wording of the claim, annotators should use their own judgment.

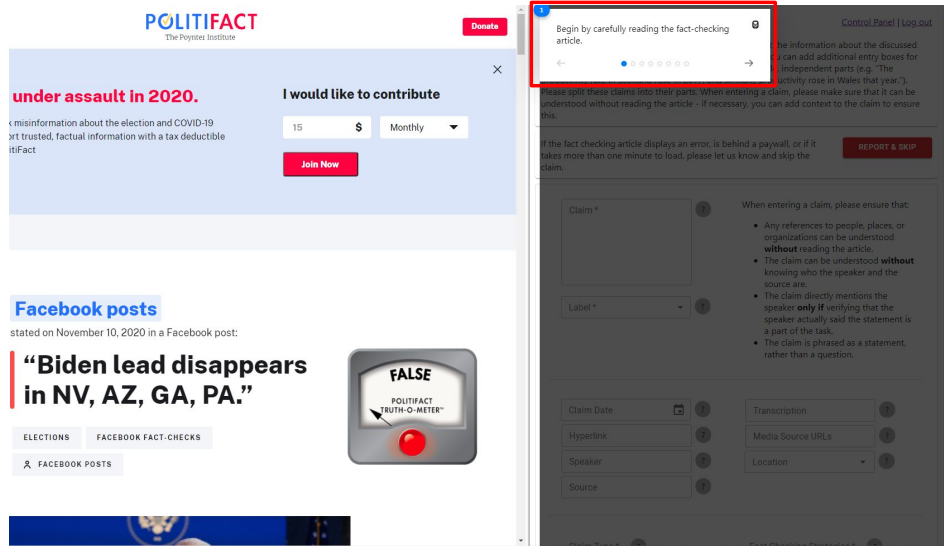


Figure 3: Interface of the tour guide.

Figure 4: Interface of claim normalization. ① The fact-checking article provided. ② Guideline of annotation for this phase. Please read it before annotating. Notice that if the article displays a 404 page or another error, or if it takes more than one minute to load, please click the **REPORT & SKIP** button. ③ Fields for the normalized claim and the corresponding label. ④ General information of the claim. ⑤ Check-boxes for selecting the type of the claim. ⑥ Check-boxes for selecting the fact-checking strategy used. ⑦ Button for adding more claims. ⑧ Buttons for submitting the current claim, going to the previous claim, and next claim.

### 3.1 Claim Splitting

Some claims consist of multiple, easily separable, independent parts (e.g. “*The productivity rate in Scotland rose in 2017, and similarly productivity rose in Wales that year.*”). The first step is to split these compound claims into individual claims. Metadata collection and normalization will then be done inde-

pendently for each individual claim, and in subsequent phases they will be treated as separate claims. When splitting a claim, it is important to ensure that each part is understandable without requiring the others as context. This can be done either by adding metadata in the appropriate field, such as the claim speaker or claim date, or through rewriting. For example, for the claim *“Amazon is doing great damage to tax paying retailers. Towns, cities and states throughout the U.S. are being hurt - many jobs being lost!”*, it should be clear what is causing job loss in the second part. A possible split would be *“Amazon is doing great damage to tax paying retailers”* and *“Towns, cities and states throughout the U.S. are being hurt by Amazon - many jobs being lost”*. That is, it is necessary to rewrite the second part by adding *Amazon* a second time in order for the second part to be understandable without context.

### 3.2 Claim Contextualization

Some claims are not complete, which means they lack adequate contextualization to be verified. For example, in the claim *“We have 21 million unemployed young men and women.”*, there are unresolved pronouns without which the claim cannot be verified (e.g. *we* refers to Nigeria, as the speaker of the claim is the presidential candidate of Nigeria). Another example is *“Israel already had 50% of its population vaccinated.”* We need to know when this claim was made to verify its veracity, as the time is crucial for this verification. For the latter, metadata is enough to resolve ambiguities; the former needs to be rewritten as *“Nigeria has 21 million unemployed young men and women.”*

Annotators are asked to contextualize claims to the original post by gathering the necessary information. Some information can be included simply as metadata, but this is not always enough – for information not captured by metadata, we ask that the claim itself is rewritten to include said information. Annotators need to follow this checklist:

1. Is the claim referring to entities which can only be identified by reading the associated fact-checking article, even if all metadata is taken into consideration? If so, add the names of the entities (e.g. *“Quotes former first lady as saying, ‘White folks are what’s wrong with America’.”* becomes *“Quotes former first lady Michelle Obama as saying, ‘White folks are what’s wrong with America’.”*).
2. Does the claim have unnecessary quotation marks, or references to a speaker (such as the word *says* in the example here)? If so, remove them (e.g. *“Says ‘Monica Lewinsky Found Dead’ in a burglary.”* becomes *“Monica Lewinsky found dead in a burglary.”*). Do NOT remove the reference to the speaker if the central problem is to determine if that person actually said the quote, e.g. in the case of quote verification.
3. Is the claim a question? If so, rephrase it as a statement (e.g. *“Did a Teamsters strike hinder aid efforts in Puerto Rico after Hurricane Maria?”* becomes *“A Teamsters strike hindered aid efforts in Puerto Rico after Hurricane Maria in 2017.”*).
4. Does the claim contain pronominal references to entities only mentioned in the fact-checking article? If so, replace the pronoun with the name of that entity. (e.g. *“We have 21 million unemployed young men and women.”*).
5. For some fact-checking articles, the title used does not properly match the fact-checked claim. Find the original claim in the article, and use that for producing the normalized version. As shown in Figure 5, the claim should be the first sentence of the article rather than the title.

Generally, try to make claims specific enough so that they can *be understood* and so that *appropriate evidence can be found* by a person who has not seen the fact-checking article.

It is important that all rewrites should be done *after* reading the article and understanding what the central problem is.

1. The claim is **supported**. The claim is supported by the arguments and evidence presented.
2. The claim is **refuted**. The claim is contradicted by the arguments and evidence presented.
3. There is **not enough information** to support or refute the claim. The evidence either directly argues that appropriate evidence cannot be found, or leaves some aspect of the claim neither supported nor refuted. We note that many fact-checking agencies mark claims as *refuted* (or similar), if supporting evidence does not exist, without giving any refuting evidence. We ask annotators to use *not enough information* for this category, regardless of the original label.
4. The claim is misleading due to **missing context**, but not explicitly refuted. This includes cherry picking, true-but-misleading claims (e.g. the claim “*Alice has never lost an election*” with evidence showing Alice has only ever run unopposed), as well as cases where conflicting or internally contradictory evidence can be found. Missing context may also be relevant if a situation has changed over time, and the claim fails to mention this (e.g. “*Alice is a strong supporter of*



*industrial subsidies*” with evidence showing that Alice currently supports the position, but in the past opposed the position).

**Important!** The label given in Phase 1 – and *only* in Phase 1 – should reflect the decision of the fact checker, not the interpretation of the annotator. In Phase 1, annotators should report the original judgment, as closely as possible, even if they disagree with it.

### 3.4 Metadata Collection

Annotators need to collect metadata through the following three steps.

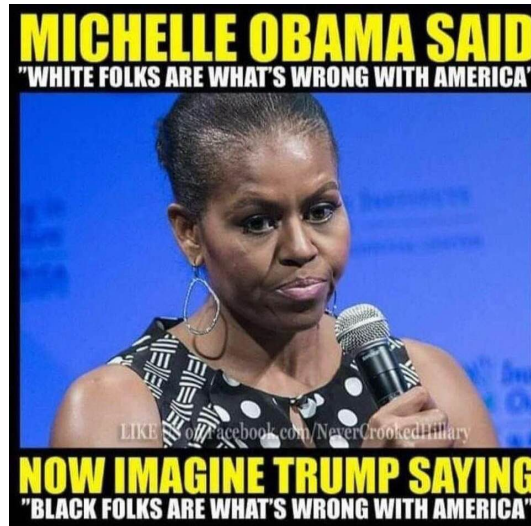


Figure 6: An example of an image claim requiring transcription.

#### 3.4.1 General Information

- A hyperlink to the original claim, if that is provided by the fact-checking site. Examples of this include Facebook posts, the original article or blog post being fact checked, and embedded video links. If the original claim has a hyperlink on the fact-checking site, but that hyperlink is dead, annotators should leave the field empty.
- The date of the original claim, regardless of whether it is necessary for verifying the claim. This date is often mentioned by the fact checker, but not in a standardized place where we could automatically retrieve it. Note that the date for the *original claim* and the *fact-checking article* (often its publication date) may be different and both stated in text. We specifically need the original claim date, as we intend to filter out evidence that appeared after that date.
- The speaker of the original claim, e.g. the person or organization who made the claim.
- The source of the original claim, e.g. the person or organization who published the claim. This is not necessarily the same as the speaker; a person might make a comment in a newspaper, in which case the person is the speaker and the newspaper is the source.
- If the original claim is or refers to an image, video, or audio file, annotators should add a link to that media file (or the page that contains the file, if the media file itself is inaccessible).
- If the original claim is an image that contains text – for example, Figure 6 shows a Facebook meme about Michelle Obama – annotators should transcribe the text that occurs in the image as metadata. In the example, it would be “Michelle Obama said white folks are what’s wrong with America.”

- If the fact-checking article is paywalled or inaccessible due to an error message, annotators should report this and skip the claim using the corresponding button.

### 3.4.2 Claim Type

The type of the claim itself, independent of the approach taken by the fact checker to verify or refute it, should be chosen from the following list. This is not a mutually exclusive choice – a claim can be speculation about a numerical fact, for example. As such, annotators should choose one *or several* from the list.

- **Speculative Claim:** The primary task is to assess whether a prediction is plausible or realistic. For example *“the price of crude oil will rise next year.”* The primary task is to assess whether the prediction is plausible or realistic.
- **Opinion Claim:** The claim is a non-factual opinion, e.g. *“cannabis should be legalized”*. This contrasts with factual claims on the same topic, such as *“legalization of cannabis has helped reduce opioid deaths.”*
- **Causal Claim:** The primary task is to assess whether one thing caused another. For example *“the price of crude oil rose because of the Suez blockage.”*
- **Numerical claim.** The primary task is to verify whether a numerical fact is true, or to verify whether a comparison between several numerical facts hold, or to determine whether a numerical trend or correlation is supported by evidence.
- **Quote Verification.** The primary task is to identify whether a quote was actually said by the supposed speaker. Claims *only* fall under this category if the quote to be verified directly figures in the claim, e.g. *“Boris Johnson told journalists ‘my favourite colour is red, because I love tomatoes’ ”*.
- **Position Statement.** The primary task is to identify whether a public figure has taken a certain position, e.g. supporting a particular policy or idea. For example, *“Edward Heath opposed privatisation”*. This also includes statements that opinions have changed, e.g. *“Edward Heath opposed privatisation before the election, but changed his mind after coming into office”*. Factual claims about the actions of people (e.g. *“Edward Heath nationalised Rolls-Royce”*) are not position statements (they are event or property claims); claims about the attitudes of people (e.g. *“Edward Heath supported the nationalisation of Rolls-Royce”*) are.
- **Event/Property Claim.** The primary task is to determine the veracity of a narrative about a particular event or series of events, or to identify whether a certain non-numerical property is true, e.g. a person attending a particular university. Some properties represent causal relationships, e.g. *“The prime minister never flies, because he has a fear of airplanes”*. In those cases, the claim should be interpreted as both a property claim and a causal claim.
- **Media Publishing Claim.** The primary task is to identify the original source for a (potentially doctored) image, video, or audio file. This covers both doctored media, and media that has been taken out of context (e.g. a politician is claimed to have shared a certain photo, and the task is to determine if they actually did). This also includes HTML-doctoring of social media posts. We will discard all claims in this category.
- **Media Analysis Claim.** The primary task is to perform complex reasoning about pieces of media, distinct from doctoring. This could for example be checking whether a geographical location is really where a video was taken, or determining whether a specific person is actually the speaker in an audio clip. The claim itself *must directly involve* media analysis; e.g. *“the speaker of these two clips is the same”*. Claims where the original source is video, but which can be understood

and verified without viewing the original source, do not fall under this category. An original video or audio file can feature as metadata in fact-checking articles, but claims are only *complex media claims* if analysis of the video or audio beyond just extracting a quote is necessary for verification.

Several claim types – speculative claims, opinion claims, media publishing claims, and media analysis claims – will not be included in later phases.

### 3.4.3 Fact-checking Strategy

After identifying the claim type, we ask annotators to classify the approach taken by the fact checker according to the article. This is independent of the claim type, as a fact-checker might take any number of approaches to a given claim. Again, one *or several* options should be chosen from the following list:

- **Written Evidence.** The fact-checking process involved finding contradicting written evidence, e.g. a news article directly refuting the claim.
- **Numerical Comparison.** The fact-checking process involved numerical comparisons, such as verifying that one number is greater than another.
- **Consultation.** The fact checkers directly reached out to relevant experts or people involved with the story, reporting new information from such sources as part of the fact-checking article.
- **Satirical Source Identification.** The fact-checking process involved identifying the source of the claim as satire, e.g. The Onion.
- **Media Source Discovery.** The fact-checking process involved finding the original source of a (potentially doctored) image, video, or soundbite.
- **Image analysis.** The fact-checking process involved image analysis, such as comparing two images.
- **Video Analysis.** The fact-checking process involved analysing video, such as identifying the people in a video clip.
- **Audio Analysis** The fact-checking process involved analysing audio, such as determining which song was played in the background of an audio recording.
- **Geolocation.** The fact-checking process involved determining the geographical location of an image or a video clip, through the comparison of landmarks to pictures from Google Streetview or similar.
- **Fact-checker Reference.** The fact-checking process involved a reference to a previous fact-check of the same claim, either by the same or a different organisation. Reasoning or evidence from the referenced article was necessary to verify the claim.

## 4 Phase 2: Question Generation and Answering

The next round of annotation aims to produce pairs of questions and answers providing evidence to verify the claim. The primary sources of evidence are the URLs linked in the fact-checking article. We also provide access to a custom search bar to retrieve evidence.

The annotator is first asked to read the claim and the associated fact-checking article (including the verdict). Following this, the annotator is asked to generate questions the answers to which are the evidence that allowed the fact checker to reach the verdict. We then ask the annotator to go through the following iterative process:

1. Ask a question to collect evidence about the claim.



**POLITIFACT**  
The Impartial Factchecker

**Facts are under assault in 2020.**  
We can't fight back misinformation about the election and COVID-19 without you. Support trusted, factual information with a tax deductible contribution to Politifact.  
[More Info](#)

**I would like to contribute**  
\$15 \$ Monthly  
[Join Now](#)

**Facebook posts**  
started on November 10, 2020 in a Facebook post:  
"Biden lead disappears in NV, AZ, GA, PA."  
**FALSE**  
POLITIFACT "WORTH A METER"  
ELECTIONS FACEBOOK FACT-CHECKS FACEBOOK POSTS

**Question Generation (1/15)**  
Please read the claim below, and the fact checking article to the left. Then, construct question-answer pairs. You can use any links in the fact checking article to provide sources for your answers. If the links in the fact checking article are not sufficient, you can also use our custom search field below. Please do not use any other search field. If you cannot find an answer to a question you ask, please select that question as "Unanswerable," and ask another question. When you have collected enough evidence to verify the claim independently of the fact checking article, please give your verdict. Please spend no more than 15 minutes on each claim.  
If the fact checking article displays an error, is behind a paywall, or if it takes more than one minute to load, please let us know and skip the claim.  
[Report & Skip](#)

**Biden lead disappears in NV, AZ, GA, PA.**  
Claim Source: Ben Ferguson  
Claim Source: Facebook  
Claim Date: 10/11/2020  
Location: US

**Question**  
Should media outlets predict Biden will win in Nevada?  
**Answers**  
One answer: ABC News, CBS News, NBC News, CNN, Fox News, Decision Desk HQ, Associated Press, Reuters, New York Times  
This answer: Extractive  
Source URL: https://web.archive.org/web/20201110120000/https://www.foxnews.com/politics/biden-lead-disappears-nv-az-ga-pa  
Web table

**When entering a question, please ensure that it:**  
• Is a well-formed sentence, rather than a search engine query.  
• Does not refer to named entities that appear only in the fact-checking article, and not in the claim, in metadata, or in previous answers.  
• Does not directly ask whether the claim holds, e.g. "Is it true that Biden?"

If you find multiple answers to your question, you can add additional answers here. Please try to rephrase the question to yield a single answer BEFORE you add additional answers.

**Submit**  
**Previous**  
**Next**

**Web Search**  
Search  
United States (US)

Figure 7: Interface of question generation. ① Guideline of annotation for this phase. Please read it before annotating. Notice that if the article displays a 404 page or another error, or if it takes more than one minute to load, please click the **REPORT & SKIP** button. ② The claim and the associated metadata. ③ Fields for the first question and its answers. Annotators can add up to 3 answers for each question if necessary. The text fields of metadata of question answer pairs are also provided. ④ Annotators can use the plus button to add as many questions as they want. Please select the label of this claim after finishing the question and answer generation. ⑤ Buttons for submitting the current claim, going to the previous claim, and next claim. ⑥ The custom search engine.

2. Attempt to answer the question. Answers should be sought from the metadata, any of the sources listed on the fact-checking article (e.g. any hyperlinks to other sites), and when that is not possible (e.g. due to the hyperlinks being dead) from the internet using our search bar.
3. If the question was answered by metadata, please write *metadata* as the source instead of a URL.
4. If the question was not answerable from any of the links on the fact-checking article, or from searching the internet, mark it as unanswerable.
5. If the claim is not verifiable using the already retrieved answers, ask a new question or rephrase an existing one (which was not answerable), proceeding from (1). Otherwise, or if more than five minutes have passed, proceed to give a verdict for the claim.
6. Assign a verdict from the list in Section 3.3 based *only* on the question-answer pairs, then proceed to the next hit.

Before proceeding to the next hit, the annotator will be shown a warning with the QA-pairs they have generated. They will also be shown their assigned label. They will be asked to confirm that the collected evidence is sufficient to assign the label they have chosen to the claim.

**Important!** Annotators should not choose a label if the retrieved evidence does not support it; for example, if the label **missing context** is chosen, there should be evidence documenting the missing

context. Labels in phase two can contradict the label of the fact-checker, if the annotator believes it is appropriate.

Metadata should only be used to generate answers, not to support a verdict without generating the appropriate question/answer pair. If, for example, the claim “*aliens made contact with earth March 3rd, 2021*” was published on September 1st, 2020, the publication date cannot be directly used to refute the claim. Instead, annotators should FIRST generate a question/answer pair – “*when was this claim made?*” “*September 1st, 2020*” – which can THEN be used to refute the claim.

## 4.1 Question Generation

To ensure the quality of the generated questions, we ask the annotators to create their questions as follows:

- Questions should be well-formed, rather than search engine queries (e.g. “where is Cambridge?” rather than “Cambridge location”).
- Questions should be standalone and understandable without any previous questions.
- Questions should be based on the version of the claim shown in the interface (i.e. the version extracted by phase one annotators), and not on the version in the fact-checking article. If an annotator believes a phase one claim has been extracted wrongly, they can correct it using the appropriate box.
- The annotators should avoid any question that directly asks whether or not the claim holds, e.g. “*is it true that [claim]*”.
- The annotators should ask all questions necessary to gather the evidence needed for the verdict, including world knowledge that might seem obvious, but could depend for example on where one is from. For example, Europeans might have better knowledge of European geography/history than Americans, and vice-versa.
- As a guiding principle, at least 2 questions should be asked. This is not a hard limit, however, and the annotators can proceed with only one question asked if they do not feel more are needed.

The following are examples used to illustrate how questions should be asked. These are based on the real claim “*the US in 2017 has the largest percentage of immigrants, almost tied now with the historical high as a percentage of immigrants living in this country*”:

- Good: What was the population of the US in 2017?
- Good: How many immigrants live in the US in 2017?
- Bad: What was the population of the US? [No time specified to find a statistic]
- Bad: What was the population there in 2017? [What does *there* refer to?]
- Bad: Is it true that the US in 2017 has the largest percentage of immigrants, almost tied now with the historical high as a percentage of immigrants living in this country? [Directly paraphrases the claim]

## 4.2 Common sense assumptions and world knowledge

As a part of the question generation process, annotators may have to make assumptions and/or use world knowledge to interpret the claim. For example, for the claim “*Shakira is Canadian*”, it may be necessary to choose what it means to be Canadian. This is expressed in how questions are formulated, e.g. “*does Shakira have Canadian citizenship?*” or “*where does Shakira live?*”. This may also involve politically

charged judgments. For example, some First Nations people are classed as Canadian by the Canadian government, but do not use that label for themselves.

In such cases, we ask annotators to follow – as closely as possible – the judgments made by the fact-checking websites. If the annotators feel that these are incomplete or misleading, they can add additional questions.

For example, for the claim “*Edward Heath opposed privatisation*”, a fact checker might provide his party manifesto as evidence. A corresponding question could then be “*what did the 1970 Conservative Party manifesto say about privatisation?*” An annotator could encounter evidence for the nationalisation of Rolls Royce during Heath’s government, which the fact-checking article did not take into account. In that case, the annotator might want to add an additional question, such as “*did Heath’s government nationalise any companies?*”. The annotators should ask *both* questions.

**Important!** As opposed to Phase 1, annotators in Phase 2 *should* use their own judgment to assign labels (although they should not ignore evidence used by the fact-checker). As such, if they disagree with the fact-checker about the label, they can select a different label.

### 4.3 Answer Generation

To find answers to questions, the annotators can rely on metadata, or on any sources linked from the factchecking site. Where these fail to produce appropriate information – either because they are not relevant to an asked question or because they refer to sources which have been taken down – we provide search functionalities as an alternative. Note that the annotators are not allowed to use the fact-checking article itself as a source, only the pages *hyper-linked* in the fact-checking article (and only when they are not from fact-checking websites).

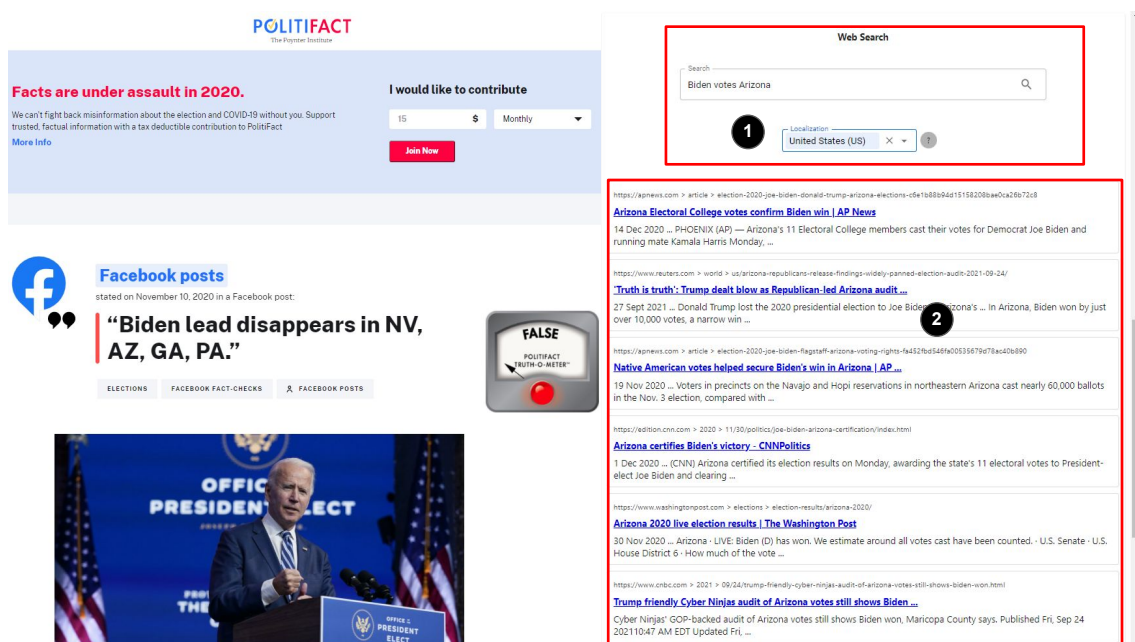


Figure 8: Interface of the search bar. ① Search bar and the location option. Annotators can change the localization of the search engine by selecting the country code here. ② Search results returned by the search engine.

Once an answer has been found, annotators can choose between the following four options to enter it:

- **Extractive:** The answer can be copied directly from the source. We ask the annotators to use their browser’s copy-paste mechanism to enter it.

- **Abstractive:** A freeform answer can be constructed based on the source, but not directly copy-pasted.
- **Boolean:** This is a special case of abstractive answers, where a yes/no is sufficient to answer the question. A second box must be used to give an explanation for the verdict grounded in the source (e.g. “yes, because...”).
- **Unanswerable:** No source can be found to answer the question.

For extractive, abstractive, and boolean answers, the annotators are also asked to copy-paste a link to the source URL they used to answer the question. Extractive answers are preferred to abstractive and boolean answers.

In some cases, annotators might find different answers from different sources. Our annotation tools allows adding additional answers, up to three. While we provide this functionality, we ask that annotators try to rephrase the question to yield a single answer before adding additional answers.

We note that if the annotators can only find a *partial* answer to a question, they can still use that. In such cases, please give the partial answer rather than marking the question as unanswerable.

Our search engine marks pages originating from known sources of misinformation and/or satire. We do not prevent annotators from using such sources, but we ask that annotators avoid them if at all possible. In the event that an annotator wishes to use information from such a source, we strongly prefer that the finds similar, corroborating information from an additional source in order to further substantiate the evidence.

While answering a question, we furthermore ask annotators to adhere to the following:

#### **Important!**

- DO NOT use any other browser window/search bar to find an answer. You MUST use the provided search bar only.
- DO NOT give a verdict for the claim until you have finished questions and answers.
- DO NOT submit answers using articles from fact-checking websites, such as politifact.com or factcheck.org, as evidence.

## **4.4 Reasoning Chain of Claims**

Annotators can build up reasoning chain across multiple questions, meaning that answers of one question can be used in the next question. For example, for the claim “*the fastest train in Japan drives at a top speed of 400 km/h*”, the first question is “What is the fastest Japanese train?”. The answer is “The fastest Japanese train is Shinkansen ALFA-X”. Based on the answer, we can further ask the second question to get more details, “What is the maximum operating speed of the Shinkansen ALFA-X”. Note that while the *generation* of the second question assumes knowledge of the answer to the first, it is *understandable* without it.

## **4.5 Confirmation**

After submitting the question/answer pairs for a claim, annotators will be presented with a confirmation screen (see Figure 9). Annotators will be shown the question/answer pairs they have entered, along with the verdict, and asked to confirm a second time that the verdict is supported by the evidence.

# **5 Phase 3: Quality Control**

Once we have collected evidence in the form of generated questions and retrieved answers, we want to provide a measure of quality. Given a claim with associated evidence, we ask a third round of annotators

Thorough hand-washing with an ordinary soap is effective in killing coronavirus (COVID-19).

Claim Label \*  
Supported

CANCEL
CONFIRM

**Confirmation (3/20)** [Control Panel](#) | [Log out](#)

Please confirm that you can infer your chosen verdict using ONLY your question-answer pairs (shown below).

Question

How does soap kill coronavirus?

Answer

The outer layer of the virus is made up of lipids, aka fat. Soap dissolves that barrier, killing the virus.

[View source](#)

Question

How does handwashing fight coronavirus?

Answer

Coronavirus can spread through tiny droplets if the infected person coughs and sneezes. Touching any surface with droplets on, and then touching your eyes, nose, or mouth, can transmit the disease.

[View source](#)

Figure 9: Before moving on to the next claim, phase two annotators will be shown a confirmation screen to make sure that their chosen verdict is correct.

to give a verdict for the claim. Crucially, the annotators at this round do not have access to the original fact-checking article, or to the claim label.

Instructions for this phase are as follows:

1. Read the claim, the metadata, and the question-answer pairs. This is the only information which should be used during this phase. Do NOT use web search to find additional information, or rely on background knowledge which an average English speaker might not have.
2. If the claim lacks context or is unreadable, flag it.
3. Flag any answers that seem incorrect, e.g. if the answer to “*how many people live in California?*” is “*thirty billion*”.
4. After reviewing the claim and the QA pairs, assign a label to the claim (see the four labels introduced in Section 3.3).
5. Finally, write a short statement justifying the verdict. If you have used any commonsense information (e.g. background knowledge which an average English speaker *is* likely to have), it should be mentioned in the justification.

We use this information to decide which claims to keep, and which to discard. Malformed QA pairs should be discarded, but the claims can be kept if there is sufficient evidence otherwise. We can potentially replace or repair these QA-pairs, and reannotate the claims.

In addition to the verdict, we as mentioned also ask annotators in Phase 3 to write a short statement justifying their verdict. Other than commonsense knowledge which an average English speaker is likely to have – e.g. “*Earth is a planet*” – there should not be any new information presented in this statement. The justification should describe how the annotators used the information present in the claim, the metadata, and the QA-pairs to reach their verdict. These justifications can be used by systems to learn to provide explanations along with verdicts.

Claim Speaker

Claim Source

Facebook

Claim Date

2020-11-10

Location

US

Justification

Many media organizations believe Biden will win in NV, AZ, GA, and PA. As such, his lead has not disappeared.

Claim Label

Refuted

☐ The claim lacks context or is otherwise impossible to understand.

Verdict Validation (1/15)

[Control Panel](#) | [Log out](#)

Please read the claim and the question-answer pairs. Then, give your verdict on the claim. Do not look at any external information; make your verdict based ONLY on the question-answer pairs. If there are any problems with a question-answer pair, please use the form to report it. Do not use the information in any question-answer pair you report to make your verdict.

Question

Which media outlets predict Biden will win in Georgia?

Potential problems with this question. Check any you think apply:
 

☐ The question is not understandable/readable
 ☐ The question is unrelated to the claim

Answer

No answer could be found.

Potential problems with this answer. Check any you think apply:
 

☐ The answer is not understandable/readable
 ☐ Answer seems wrong, but is supported by the source
 ☐ Answer seems wrong, and is not supported by the source
 ☐ I believe the source may be biased

Question

Which media outlets predict Biden will win in Arizona?

Potential problems with this question. Check any you think apply:
 

☐ The question is not understandable/readable
 ☐ The question is unrelated to the claim

Answer

Fox News

Potential problems with this answer. Check any you think apply:
 

☐ The answer is not understandable/readable
 ☐ Answer seems wrong, but is supported by the source
 ☐ Answer seems wrong, and is not supported by the source
 ☐ I believe the source may be biased

Answer

Associated Press

Potential problems with this answer. Check any you think apply:
 

☐ The answer is not understandable/readable
 ☐ Answer seems wrong, but is supported by the source
 ☐ Answer seems wrong, and is not supported by the source
 ☐ I believe the source may be biased

PREVIOUS

SUBMIT

NEXT

Figure 10: Interface of quality control. ① Text field for entering the justification. ② Label of the claim and the checkbox of unreadable. Notice that once the unreadable option is selected, annotators do not need to select the label for the claim. ③ The question corresponds to the current claim. Here we have two question-answer pairs. If the annotator think the there exist potential problems with this question, check any options applied. ④ The answers corresponds to the question on the left. If the annotator think the there exist potential problems with the answer, check any options applied. ⑤ Buttons for submitting the current claim, going to the previous claim, and next claim.