


Guideline for AVeriTeC-IT Annotators

May 16, 2025

1 Introduction

We aim to construct a dataset for automated multimodal fact-checking, as multimodal misinformation spreads faster and seems to be more convincing than text-only one. Specifically, we focus on fact-checking image-text claims (i.e., textual claims associated images and both texts and images are indispensable for fact-checking). The construction of the dataset follows three guiding principles: 1) we focus on the image-text claims, the verification of which must involve both texts and images; 2) we intend to decompose the evidence retrieval process into multiple steps, annotating each individual step as a question-answer pair (as illustrated in Figure 1¹); 3) our dataset will be constructed from real-world image-text claims previously checked by journalistic organizations, rather than the artificially created claims used in prior work (e.g., NewsCLippings Luo et al. (2021)).

Claim: US Congresswoman Ilhan Omar at a training camp of a Somali warlord



Q1: What did the image show?
A1: It showed a woman recruit of the Somali Army at a military training campus at Halane, Mogadishu

Q2: When was the image taken?
A2: The image was shot on February 25, 1978.

Q3: When was Omar born?
A3: She was born in 1982.

Verdict: Refuted (Misuse of Images: Out-of-context)
Justification: The claim refers an image to Ilhan Omar. However, the image was taken on 1978, while Omar was born four years later in 1982.

Figure 1: An example of an image-text claim and its question-answer pairs for verification.

Verifying an image-text claim necessitates both the text and the image content of the claim. For example, to verify the claim in Figure 1, it requires checking if the textual part of the claim aligns with the image. Decomposing claim verification into question generation and answering allows us to break complex real-world image-text claims to their components, simplifying the task. Specifically, verifying the claim in

¹The original fact-checking article for the example: <https://factcheck.afp.com/doc.afp.com.34X68NX>

Figure 1 requires information about the image and knowledge about the date of birth of Ilhan Omar. Three separate questions need to be asked in order to reach a verdict (i.e., whether the claim is *supported* or *refuted*).

By decomposing the evidence retrieval process in this way, we attempt to emulate the work of real-world fact-checkers. It is a natural way for systems to justify their verdicts and explain their reasoning to users. In addition, we annotate claims with a final justification, providing a textual explanation of how to combine the retrieved answers to reach a verdict.

The annotation consists of the following three phases:

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

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.

2 Interface

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! If you are logged into multiple sessions using the same account, the annotation tool may lose the data you enter.
- If the logging in has an error message: *Error: Request failed with status code 401*, it is probably because your previous session has expired. You need to log out first and log in again.
- If you think the images in the interface are showing inappropriately (e.g., some part of the image is hidden), you can try to adjust the scale or size of your web pages.



Figure 2: Interface of the control panel. (1) Button for the guideline. (2) Button for changing the password. (3) Button for logout. (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. (5) Start the annotation for this phase. Here is Phase 1 Claim Normalization.

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 for that phase. Please read the hints provided by the tour guide carefully before the annotation.

3 Phase 1: Claim Normalization

In this phase, annotators produce a normalized version of each claim, choose the claim modality types (i.e., what modalities of the claim are involved during fact-checking) and collect metadata about claims. Below is an overview of the claim normalization task (details can be found in the subsequent sections).

1. First, annotators should read the fact-checking article and identify which claims are being investigated.
2. If a fact-checking article contains more than one claims or claims with multiple parts, the annotators should first split claims into independent parts (see Section 3.1).
3. Annotators should provide a normalized version of texts in a claim. The normalization means the extracted claim, specifically the textual part of claim, can be understood without the context of the fact-checking article (e.g., the text refers to entities in the fact-checking article) and is not ambiguous or vague.
4. Generally, we prefer claims to be as close as possible to their original form (i.e. the form originally said, not necessarily the form used in the fact-checking article). As such, contextualization should be done only when necessary, following the checklist in Section 3.2.
5. Annotators should choose the modality type for each claim (see Section 3.3). Except *image-text* claims, claims of other modality types will be filtered out (i.e., will not be passed to annotation steps below). For some *image-text* claims, the images are not involved during fact-checking. We provide an option for annotators to select if the images are used during the verification process. Note, the verification process is the one used in the fact-checking article, rather than annotators' own judgement.
6. Annotators should upload images as part of a normalized claim for image-text claims. The uploaded image should be as close as possible to its original form but potentially removing added content by fact-checkers (e.g., a stamp of "hoax" added to the image) (see Section 3.4).
7. Annotators should extract the verdict of the claim in the article and translate it as closely as possible to one of our four labels - *supported*, *refuted*, *not enough evidence*, and *conflicting evidence/cherry picking* (see Section 3.5). In phase 1, annotators should give their own judgments, however, they should match as closely as possible the judgments given by the fact-checking articles.
8. For *refuted* image-text claims, annotators should provide reasons why they are refuted. The reason could be *textual refuted*, *misuse of image* or *others* (see Section 3.6). The reasons are not mutually exclusive, as a claim could be refuted because of misuse of both text and image.
9. For claims misusing images, annotators should state the reason(s) for misuse according to the fact-checking article as closely as possible to our three labels - *out-of-context*, *image manipulation*, and *other misuse types* (see Section 3.6). These are not mutually exclusive, and more than one misuse type can be chosen.
10. Claims will have associated metadata (e.g., the date the original claim was made, or the name of the person who made it). Annotators should identify and extract this metadata from the article (see Section 3.7).
11. Annotators should identify the type of each claim, choosing from the options described in Section 3.8. These are not mutually exclusive, and more than one claim type can be chosen.
12. Annotators should identify the strategies used in the fact-checking article to verify each claim, choosing from the options described in Section 3.9. These are not mutually exclusive, and more than one claim type can be chosen.

Important! Annotators' decisions (e.g., what modalities are involved for fact-checking or the labels of a claim) should match as closely as possible the fact-checking articles in Phase 1.

3.1 Claim Splitting

Some fact-checking articles have multiple claims, meanwhile, some claims are compound and have been divided into multiple claims by fact-checkers. For instance, the compound claim, “*The man in the image was struck, ‘roasted’ by thunderstorm while receiving phone call.*”² was divided into three claims by the fact-checker: 1) *The man in the image was receiving phone call in the rain;* 2) *Thunder -lightning-struck and the man was burnt (roasted)* and 3) *Receiving or making calls during thunderstorms can get one burnt (roasted).* We intend to conduct claim splitting to separate these claims into their individual and independent parts. The claim splitting serves as the first step in the phase one, as all following annotations should be done on individual 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 “*One Rafale jet crashed during training. Two pilots died.*”³, it should be clear what was the cause of death of the two pilots in the second part. A possible split would be “*One Rafale jet crashed during training.*” and “*Two pilots died in the Rafale jet crash.*”. That is, it is necessary to rewrite the second part by adding the cause of death for the second part to be understandable without context.

3.2 Claim Contextualization

Some claims, specifically their textual parts, are not complete, which means they lack adequate contextualization to be verified. For example, in the textual part of a claim “He threatened people of Gujarat India in the speech if they didn’t support him.”, there are unresolved pronouns without which the claim cannot be verified (e.g. he refers to Arvind Kejriwal, the speaker of the speech). Another example is the claim⁴, the textual part of which is “*The photo was taken today at Lal Chowk.*”. 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 “*Arvind Kejriwal threatened people of Gujarat India in the speech if they didn’t support him.*”.

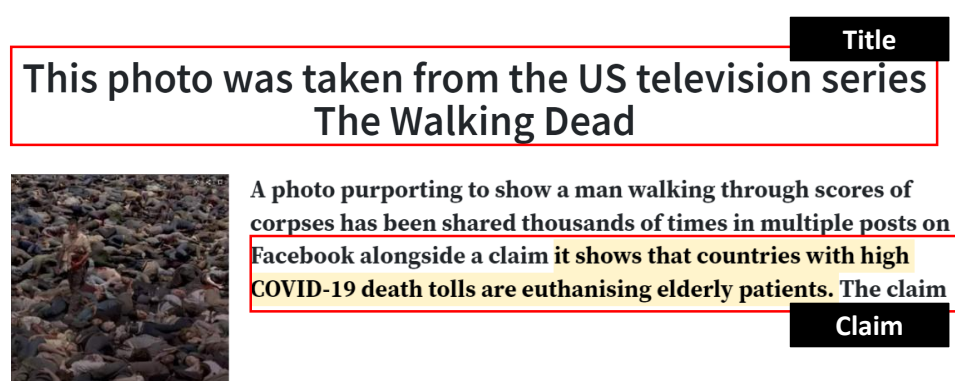


Figure 3: An example of locating the claim.

Annotators are asked to contextualize textual parts of claims to the original post by gathering the necessary information. Some information can be included simply as metadata, but this is not always enough

²<https://web.archive.org/web/20210101200024/https://www.icirnigeria.org/fact-check-was-this-man-struck-roasted-by-thunderstorm-while-receiving-phone-call/>

³<https://web.archive.org/web/20210225083155/https://newsmobile.in/articles/2020/09/12/fact-check-as-india-inducts-rafale-jets-old-pictures-of-air-crash-falsely-shared-as-rafale-crash/>

⁴<https://factly.in/an-old-photo-is-being-falsely-shared-as-a-photo-taken-today-at-lal-chowk-srinagar/>

– 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 textual part of 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. “*Nigerian leader tested positive for coronavirus*.” becomes “*Nigerian leader, Muhammadu Buhari, tested positive for coronavirus*”) ⁵.
2. Does the textual part of the claim have unnecessary quotation marks or clause such as the word “says” in the example here)? If so, remove them (e.g. “**Says** “*Joe Biden said voting by email is a way to ‘fundamentally change this country’.*”” becomes “*Joe Biden said voting by email is a way to ‘fundamentally change this country’.*” after removing “says” and the unnecessary quotes). 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 textual part of the claim a question? If so, rephrase it as a statement (e.g. “*Are these viral images of cash and gold seizures from Stalin daughter’s house?*” becomes “*These viral images showed the money confiscated from the house of Stalin’s daughter.*”) ⁶.
4. Does the textual part of the claim contain pronominal references to entities or sentences only mentioned in the fact-checking article? If so, replace the pronoun with the name of that entity. (e.g. “*Actor Jaaved Jaaferi tweeted this.*” becomes “*Actor Jaaved Jaaferi tweeted: It is not necessary that all Muslims putting saliva on fruits and vegetables are corona positive. Still, some Hindu customers are spreading hatred on Muslim vendors by boycotting them. Why are you so intolerant?*”) ⁷.
5. For some fact-checking articles, the title used does not properly match the texts in fact-checked claim. Find the original claim in the article, and use that for producing the normalized version. As shown in Figure 3, the textual part of the claim should be the first sentence of the article rather than the title.
6. Is the textual part of the claim too vague to be investigated through the use of evidence, and does the fact-checking article investigate a more specific version of the claim? If so, use the text investigated in the fact-checking article (e.g. “*Towns, cities and states throughout the U.S. are being hurt by Amazon*” might become “*Towns, cities and states throughout the U.S. are losing state tax revenue because of Amazon*”) ⁸.

Textual claims associated with media (e.g., images and videos) may imply (according to the fact-checking article) the alignment between their texts and their media. For instance, the claim in Figure 1 assumes the image is associated with the textual part of the claim. As people read the parts of multimodal claims together (e.g., an image paired with its textual part), the explicit mention of alignment (e.g., “*the image shows ...*”) should not be added in normalisation. Annotators should just make the normalized textual parts of such multimodal claims as close as possible to their original form, as long as there is no ambiguity.

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.

⁵<https://web.archive.org/web/20210118081120/https://factcheck.afp.com/cnn-broadcast-has-been-doctored-nigerian-leader-did-not-test-positive-coronavirus>

⁶<https://web.archive.org/web/20210721131607/https://newsmeter.in/fact-check/fact-check-are-these-viral-images-of-cash-and-gold-seizures-from-stalin-daughters-house-676384>

⁷<https://web.archive.org/web/20201128191527/http://newsmobile.in/articles/2020/04/19/jaaved-jaaferi-did-not-tweet-this-heres-the-fact-check/>

⁸<https://www.nytimes.com/2017/08/16/us/politics/trump-amazon-taxes.html>

3.3 Modality Type

The claim modality type refers to modalities involved in a claim. Below we provide six claim modality types, which are mutually exclusive - a claim can be only assigned to one claim modality type.



Figure 4: Two examples of different formats of text-only claims.

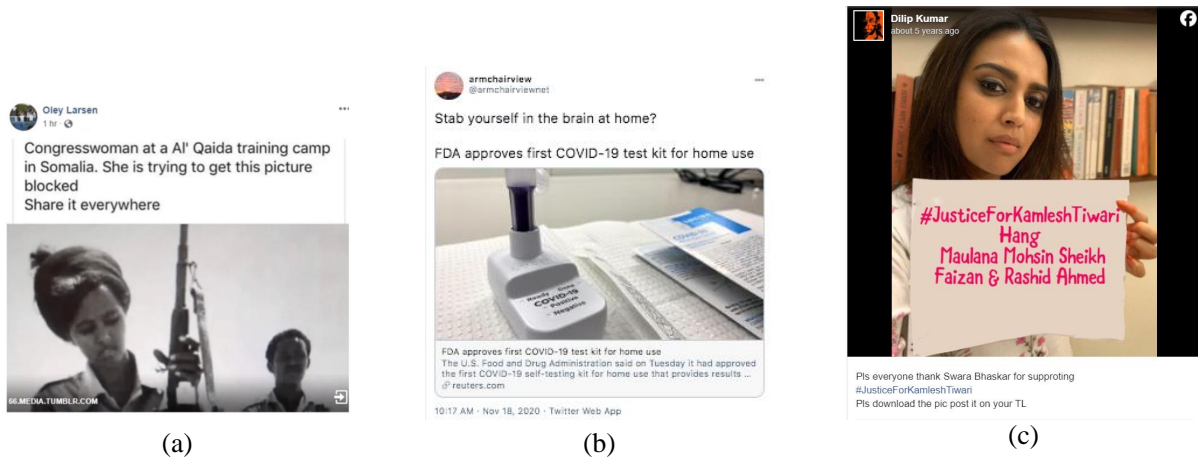


Figure 5: Three examples of image-text claims.

- **Text-Only:** The text-only type refers to claims which are purely textual. Some claims are clearly text-only and no images presented related to the claim, as illustrated in Figure 4(a). Some pure text-based claims are embedded in images. However, with OCR detection, the claim can be extracted and represented with the text-only format, as shown in Figure 4(b).
- **Image-Text:** The image-text claims are textual claims associated with images, as shown in Figure 5. The images may be relevant to the claims (e.g., as evidence to support the claim) or used for increasing the credibility of claims.
- **Video-Text:** The video-text claims are textual claims associated with videos.
- **Audio-Text:** The audio-text claims are textual claims associated with audio.
- **Image-Only:** This type of claims only involves images (e.g., deepfake or doctored image). Besides, there is no text associated with the image.

- **Other:** There could be other types of claim modality types, such as audio-only (deep fake audio of Biden ⁹) or video-only (e.g., Nvidia’s deep fake video of CEO Jensen Huang at GTC). If annotators find the claim not falling into the five types above, they should choose *other* and also need to provide a justification why they think the claim does not belong to the claim modality types above.

Only *Image-Text* claims will be considered in the following steps and later phases. Some image-text claims involve both texts and media, whereas, the media (i.e., image) is not used during fact-checking. For instance, the claim in Figure 5(b), though with an image, the image is not involved during fact-checking, because the image was not considered to be either part of the claim or increasing its credibility ¹⁰. Here, we provide an additional option for annotators to select whether the images of image-text claims are used during fact-checking. For the examples of Figure 5(a) and (c), annotators should select *yes* as the images of the two claims are used for fact-checking. Specifically, verifying the claim in Figure 5(a) ¹¹ requires checking if the woman in the image is the US congresswoman, Ilhan Omar (checking whether the textual claim aligns with the image). The example in Figure 5(b) ¹² needs to check if the image is manipulated in order to support the textual part of the claim. To decide whether the image is relevant to the fact-checking process, the annotators should follow the fact-checking article, not their own judgement.



Figure 6: (a) is an example of an image in need of normalization in an image-text claim. (b) is the desired normalized image.

3.4 Image Uploading

If a claim is identified as an *image-text* claim (according to the definition in Section 3.3), annotators should upload the image part of the claim as well. Annotators should upload all images in the original claim. We would like annotators to keep as closely as possible to the original images in the claim. We recommend annotators to open the link for the original claim in the fact-checking article and save images in the claim with right-clicks on images. Alternatively, annotators extract images from the screenshots provided by the fact-checking article. If the screenshot contains more than the images of a claim (e.g., a screenshot of an online post), annotators should use snipping tools to crop the image part from the screenshot. Annotators should not upload a screenshot of the claim but only the image part in the claim. Fact-checkers may sometime edit the image in the claim (e.g., adding “*hoax*” or “*fake*” stamps to the image, as shown in Figure 6(a). Regarding to these cases, annotators should upload the original version

⁹<https://www.wired.com/story/biden-robocall-deepfake-danger/>

¹⁰The fact-checking article to the example: <https://web.archive.org/web/20210227031424/https://misbar.com/en/factcheck/2021/01/08/fda-approves-at-home-covid-test>

¹¹The fact-checking article for the example: <https://web.archive.org/web/20210802100249/https://www.altnews.in/no-this-is-not-us-congresswoman-ilhan-omar-at-a-training-camp-of-a-somali-warlord/>

¹²<https://www.indiatoday.in/fact-check/story/fact-check-swara-bhaskar-s-placard-seeking-justice-for-kamlesh-tiwari-is-photoshopped-1612609-2019-10-24>

of the image of the claim (e.g., via the link for the original claim in the fact-checking article), rather than the edited image by the fact-checker from the fact-checking article.

In some cases, the image in a claim has been edited already in the original post. We would like the annotators to keep the original version of the image in the claim rather than seeking for other versions.

3.5 Verdict of a Claim

We ask annotators to produce a label for an image-text claim relying only on the information on the fact-checking site (and assuming that everything reported there is accurate). Note, the verdict is for the whole claim (i.e., taking both the image and text into account). For the dataset we are creating, we will be using four labels:

1. The claim is **supported**. The claim is supported by the evidence presented.
2. The claim is **refuted**. The claim (either the text or the image part) is contradicted by the evidence presented. In some cases, the textual part of an image-text claim is factually correct while the image is misused, the whole claim is still regarded as refuted. For instance, a real claim ¹³ in our dataset associated the textual claim “*Police in Karnataka are asking lockdown violators to clean up public places as punishment.*” with an image showing people cleaning a lake. The textual part of the claim is true that police in Karnataka were punishing lockdown violators by making them clean public places. However, the image was misused in a wrong context as it showed techies who had voluntarily come to clean Hulimavu Lake in Bengaluru. Considering the whole claim, it should still be labeled as *refuted*. More elaborations about how images could be misused can be found in Section 3.6.
3. There is **not enough evidence** (NEE) 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 evidence for this category, regardless of the original label. In situations where evidence can be found that the claim is unlikely to be supported, even if the evidence is not conclusive, annotators may use refuted; here, annotators should use their own judgment. We give a few examples in Section 3.5.1. If a claim is not verified, annotators should also pick this category.
4. The claim is misleading due to **conflicting evidence/cherry-picking**, but not explicitly refuted. This includes cherry-picking¹⁴ (e.g. a claim ¹⁵ “*The image shows an illustration of how a sample for a COVID-19 test is taken*”, however not all tests are done in that way), true-but-misleading (e.g. a claim ¹⁶ “*Photo shows an astronaut statue in a church built in the 1600s.* ”. The church was built in the 1600s whereas the astronaut statue was recently built), as well as cases where conflicting or internally contradictory evidence can be found. Conflicting evidence may also be relevant if a situation has recently changed, and the claim fails to mention this (e.g. “*Alice is a strong supporter of industrial subsidies*” with evidence showing that Alice currently supports industrial subsidies, but in the past opposed industrial subsidies). We note that if a claim covers a period of time, and evidence refutes the claim at some time points but not others, the whole text part is still refuted – for example, “*Alice has always been a strong supporter of industrial subsidies*” or “*Alice has never been a strong supporter of industrial subsidies*”.

¹³<https://web.archive.org/web/20201015230605/https://www.indiatoday.in/fact-check/story/truth-behind-lockdown-violators-cleaning-lake-karnataka-1666510-2020-04-13>

¹⁴<https://en.wikipedia.org/wiki/Cherry-picking>

¹⁵<https://web.archive.org/web/20200709105605/https://fullfact.org/online/nose-swab-picture/>

¹⁶<https://web.archive.org/web/20210427192059/https://factly.in/this-astronaut-sculpture-was-added-to-the-centuries-old-church-during-the-renovation-works-in-the-1990s/>

Despite the claim splitting sub-task, some texts may contain multiple parts that are too interconnected to split. This could for example be a claim with the text part like “A woman jumped off a building because she suffered from COVID-19.”. In such cases, parts of the claim may have different truth values. We discuss a few cases below:

- The text part is implicature, i.e. “*X happens because Y*” or “*X leads to Y*”. In this case, annotators should find a label for the causal implication, and not for either of the component text part.
- The text part has two components, where one is *refuted* and the other is *not enough evidence*. In this case, the entire text part should be labeled *refuted*.
- The text part has two components, where one is *supported* and the other is *not enough evidence*. In this case, the entire text part should be labeled *not enough evidence*.

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.

3.5.1 Deciding Between Refuted and NEE

As mentioned, the line between refuted and not enough evidence requires annotators to rely on their own judgment in cases where refuting evidence cannot be directly found, but the text part is extremely unlikely. As a guiding principle, if annotators doubt the truth value of the text part – given the presented evidence and/or lack of evidence – not enough evidence should be chosen. Below, we give several examples:

- A textual claim “*The 392-year-old shark was recently discovered in the Arctic Ocean.*” associated with an image showing a swimming shark in the ocean ¹⁷. The photo was an authenticated image of a Greenland shark in the Arctic Ocean from a 2016 study on the sharks. However, existing studies can only prove that the shark was at least 272 years old and could have been as much 512 years old. No evidence can prove the specific age of the shark. As such, annotators should select not enough evidence as the label.
- A textual claim “*Government ministers broke their own Covid rules at a recent ‘Cobra’ meeting.*” associated with an image showing government ministers violating the social distance rule ¹⁸. The picture is genuine, not altered and indeed shows ministers in a meeting without masks. However, it is hard to tell from photographs about the distance as photographs can make people seem to be closer with certain angles and equipment. Therefore, annotators should choose not enough evidence rather than refuted.
- “*Shakira is Canadian.*” Evidence can be found that Shakira is usually described as Colombian, was born in Colombia, and holds Colombian citizenship. Furthermore, evidence shows she now resides in Spain. As no evidence of any connection to Canada can be found despite the wealth of information available about her, it is extremely unlikely that she is secretly Canadian; therefore, the annotator can select refuted as the label.

3.6 Reasons for Refuted

For all *refuted* image-text claims, we ask annotators to provide more fine-grained reasons for refuting, in a hierarchical manner. To begin with, considering there are two modalities involved, annotators should identify which modality is the reason the claim is refuted. We consider three labels for which modality leads to refuting:

¹⁷<https://web.archive.org/web/20200226041229/https://www.politifact.com/factchecks/2020/jan/28/facebook-posts/viral-photo-does-show-longest-living-shark-species/>

¹⁸<https://web.archive.org/web/20201128221530/https://fullfact.org/health/parliament-social-distancing/>

1. **Textual refuted:** The textual part of an image-text claim is factually wrong and the refutation is not related to the image (e.g., the problem is not that the image is misinterpreted in a wrong context or manipulated). For instance, the image-text claim ¹⁹ states “*This is the picture of Dr Vidhi from Ahmedabad who succumbed to Covid-19.*”. If according to the fact-checking article the image is of Dr Vidhi, but Dr Vidhi *did not succumb to Covid-19*, and the evidence was not found through the image(s) of the claim, *Textual refuted* should be chosen as the reason for refuting the claim.
2. **Misuse of images:** The reasons for refuting an image-text claim is related to the images of the claim, and finding evidence for refuting it involves the images. An image could be misused because of misinterpretation of the image content. In some cases, the textual part of the claim is true, as shown in the example mentioned in Section 3.5: “*police in Karnataka are asking lockdown violators to clean up public places as punishment*”. The reason for refuting the claim is the (mis-)use of the image out of its context. There are also cases that the textual part of the claim maybe false together with the misuse of the image, as illustrated in Figure 1. Specifically, the textual part of the claim “*US Congresswoman Ilhan Omar at a training camp of a Somali warlord*” could be possibly wrong. But as long as the reason for refuting the textual part of the claim provided in the fact-checking article – in this case, mis-alignment between the image and the text as the image backing up the claim is not of Ilhan Omar – is only the use of images, annotators should choose *Misuse of images* rather than *Textual refuted*. If the article also provided other evidence found not through the images of the claim to refute the textual part of the claim (e.g., some articles have proven that Ilhan Omar has never been at a training camp of a Somali warlord), then the annotators should choose both *Textual refuted* and *Misuse of images*. Another case of image misuse could be digital editing of images (e.g., image content manipulations and image metadata doctoring). Similarly, if the only reason for refuting the claim (or the textual part of the claim) is related to image manipulations, annotators should choose *Misuse of images* rather than *Textual refuted*.
3. **Others:** If the annotators find the categories above cannot describe the reasons for refuting properly, they can select *Others* and give explanations about why the reason for refuting the claim does not fall into the two categories above.

The three options are not mutually exclusive and a claim can be refuting due to multiple reasons.

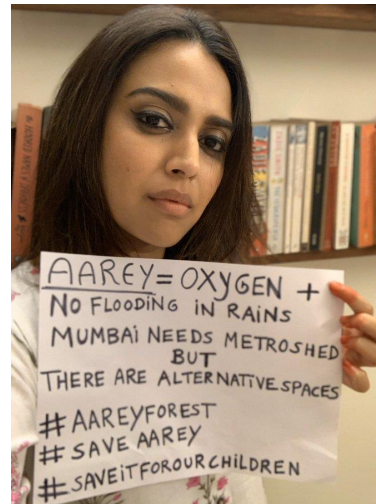
Secondly, for the cases related to *misuse of images*, annotators should further identify how images are misused according to the fact-checking article. We provide three types for misuse of images, which are not mutually exclusive - a claim can be assigned multiple types:

1. **Out-of-Context:** An image re-purposed to support other narratives by misrepresenting its context, and/or elements. For instance, the claim in Figure 1 referred to the woman in the image as the congresswoman, Ilhan Omar, in Somalia. However, the image was used out-of-context as the woman in the image was not Ilhan Omar (the photo was taken before Ilhan was born). A typical case of out-of-context misuse of image is referring an old image to a recent event (e.g., using a picture taken before COVID-19 to support the textual claim that government ministers did not wear masks during a meeting).
2. **Image manipulation:** Digital manipulation about an image. For example, the claim in Figure 7(a) uses an image with manipulated content for supporting the textual part of the claim. Figure 7(b) shows the original image. Doctoring the metadata of images (e.g., the date and time for taking the image) is also a kind of commonly seen image manipulation.
3. **Others:** If the reason for misuse does not fall in the three categories above, annotators should choose *others* and also need to provide a justification why they think the type of image misuse does not belong to the types above.

¹⁹<https://www.indiatoday.in/fact-check/story/fact-check-actor-s-picture-shared-on-social-media-as-doc-who-died-of-covid-19-1722520-2020-09-16>



(a)



(b)

Figure 7: (a) is an example of an image-text claim with image manipulation. (b) is the original image.

Note, the selection of fine-grained reasons for refuting should be based on the fact-checking articles. We will exclude examples annotated as *image manipulation* misuse type in the later phases.

3.7 Metadata Collection

Annotators need to collect metadata regarding the following aspects:

1. **Hyperlink:** A hyperlink to the original claim, if that is provided by the fact-checking site. Examples of this include Facebook posts and the original article or blog post being fact-checked. If the original claim has a hyperlink on the fact-checking site, but that hyperlink is dead, annotators should leave the field empty.
2. **Claim Date:** 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 may be stated in the text. We specifically need the original claim date, as we intend to filter out evidence that appeared after that date. The dates can often be found in the original online posts or publishing dates of articles. If multiple dates are mentioned (e.g., multiple posts about the same claim), the earliest should be used. If an imprecise date is given (e.g. February 2017), the earliest possible interpretation should be used (i.e. February 1st, 2017).
3. **Speaker:** The speaker of the original claim, e.g. the person or organization who made the claim.
4. **Source:** The source of the original claim, e.g. the platform or the journal where the claim was published. If a claim was published on Twitter, Twitter is the source of the claim. If the claim to be checked is an article, the source of the claim is publisher then.
5. **Transcription:** If an image which has texts on it, is used for representing a claim (as illustrate in Figure 7), annotators should provide a transcript of texts on the image (recognized texts on the image). For instance, the transcription for the example in Figure 7(a) should be “#JusticeForKamleshTiwarei ...”. Note, some claims are meme-based. These claims are in need of transcription (i.e., meme text extraction).

6. **Media URL:** 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). For instance, considering the claim in Figure 1, the URLs for the images in the post should be extracted.
7. **Location:** The location (i.e., country) that is most relevant to the original claim.

3.8 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 (due to limitation of space, we use text-only claims or omit the media part in examples below for illustration purpose).

- **Speculative Claim.** The primary task is to assess whether a *prediction* is plausible or realistic. For example “*Climate change will lead to many extreme weather events.*”
- **Opinion Claim.** The claim is a non-factual opinion, e.g. “*US should care more for native-born Americans.*”. This contrasts with factual claims on the same topic, such as “*US economic figures show US should care more for native-born Americans as employment gains are benefiting immigrants more than native-born Americans.*”
- **Causal Claim:** The primary task is to assess whether one thing caused another. For example “*the woman jumped off the building because of suffering from COVID-19.*”.
- **Numerical claim.** The primary task is to verify whether a numerical fact is true (e.g., “*China responsible for 90 percent of the world’s carbon emissions.*”), or to verify whether a comparison between several numerical facts holds (e.g., “*Sulphur dioxide levels rise in Wuhan post coronavirus outbreak.*”), or to determine whether a numerical trend or correlation is supported by the evidence (e.g., “*Covid-19 vaccines have a positive correlation with death rate.*”).
- **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. “*Joe Biden said voting by email is a way to ‘fundamentally change this country’*”. The verification is on whether the exact quote was said (e.g., “*fundamentally change this country*”) by Biden.
- **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, “*Ex-South Korea president Moon Jae-in supported his wife’s India visit*”. This also includes statements where positions have changed, e.g. “*Ex-South Korea president Moon Jae-in supported his wife’s India visit, but changed his mind when having a TV interview*”. Factual claims about the actions of people (e.g. “*Ex-South Korea president Moon Jae-in published an online post for supporting his wife’s India visit*”) are **not** position statements (they are event or property claims); claims about the attitudes of people are. As opposed to quote verification, the position-taker does not need to have used the *exact* words of the claim; appropriate evidence would for example be “*Moon Jae-in said in his speech that his wife’s tourist trip to India shouldn’t be criticised.*”.
- **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. “*tourist trip shouldn’t be criticised.*”. 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 if a piece of media is indeed from a certain source (e.g., published by a certain person or a certain publisher). The claim itself *must explicitly* mention the need to identify the original source for media (e.g., a politician is claimed to have shared a certain photo and the task is to determine if they actually did).
- **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 an image was taken, whether an image was taken on a certain date, whether a certain person is really the person in the image or comparing the content of two images to see if the original image has been manipulated. The claim itself *must directly involve* media analysis; e.g. “the gestures of the person in two images are different”.

Speculative claims and opinion claims will not be included in later phases.

3.9 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. This is not a mutually exclusive choice - one or several options should be chosen from the following list:

- **Written Evidence.** The fact-checking process involved finding contradicting or supporting written evidence, e.g., a news article directly refuting or supporting 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.
- **Reverse Image Search.** To find background information about images, fact-checkers usually conduct reverse image search by passing images into search engines. Usually, fact-checkers mention this strategy explicitly if it is exploited.
- **Keyword Search.** To obtain more related information about media, the fact-checking process involves searching with keywords related to the media. Usually, fact-checkers mention this strategy explicitly if it is exploited.
- **Media Source Discovery.** In some cases, fact-checkers find the source of the original media for fact-checking. If the way for tracing the source is not *reverse image search* nor *keyword search*, annotators should select this option.
- **Image analysis.** The fact-checking process involved image analysis, such as comparing two images. For instance, to check the claim in Figure 7(a), after finding the origin of the media, comparison between two images is conducted to check if the image has been manipulated.
- **Geolocation.** The fact-checking process involved determining the geographical location of an image, through the comparison of landmarks to pictures from Google Streetview or similar. Note, if the geolocation is retrieved via finding the source of the media (e.g., the retrieved web page of the original image by *Reverse Image Search* tells the geolocation of the image), it does not fall in this category.

- **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 organization. Reasoning or evidence from the referenced article was necessary to verify the claim.

Claims *only* labeled as solved through Fact-checker Reference will not be included in later phases

4 Phase 2: Question Generation and Answering

This round of annotation aims to produce pairs of questions and answers providing evidence to verify the claim. Specifically, in this phase, we *only* involve *image-text* claims the images of which are used during the fact-checking process and which are not under the *image manipulation* misuse type. The primary sources of evidence are the URLs linked in the fact-checking article. We also provide access to a custom search bar (for both text-based search and image-based search) to retrieve evidence.

In this phase, annotators are asked to conduct a question generation and answering task. Below is a quick overview of the task (details can be found in the subsequent sections).

1. The annotator should first read the claim (the textual part and the associated image(s)), metadata provided by the annotation in Phase 1 and the associated fact-checking article ²⁰ (including the verdict). We note that since Phase 1 annotators may decompose a claim into several subclaims, in some cases not all sections of the fact-checking article will be relevant.
2. If an annotator believes a claim from Phase 1 has been extracted wrongly, they can correct it. For the textual part, they can correct using the text box. For the images, annotators can upload a new set of images if they do not agree with uploaded images by the previous annotator. This should not be necessary for most claims, but adds an extra round of quality control. Guidance on correcting claims along with examples can be found in Section 4.1.
3. The task is to generate questions and answers about the claim such that a verdict can be given about the claim using them alone, without knowledge of the fact-checking article. The sources and strategies used in the fact-checking article can serve as inspiration for questions and evidence for answers, but the fact-checking article should not be directly referenced as a source. We recommend constructing question-answer pairs iteratively, one at a time. That is, annotators should ask a question and attempt to answer it, and only then proceed to the next question. Details for question generation can be found in Section 4.2.
4. For each asked question, annotators should describe what the question is concerned with (i.e., the type of the question) and try to translate it as closely as possible to our provided labels: *text-related*, *image-related*, *metadata-related* and *commonsense-related* (see Section 4.3). These options are not mutually exclusive, and one question can be associated with multiple types. For image-related questions, annotators should select the images concerned from the image candidate list, which is initialized with the uploaded images from Phase One (or revised set of images if annotators correct the image list from Phase One). The image candidate list will be updated accordingly during the QA generation process by adding answers in the format of images. Annotators are prohibited from asking questions about images not in the candidate list.
5. Annotators should provide the method for answering each question (i.e., finding evidence for each question), via *Text Search*, *Image Search*, *Image Analysis* and *Metadata* (See Section 4.4). For both *Image Search* and *Image Analysis*, annotators should also identify related images from the image candidate list.

²⁰WARNING: For persistence, we have stored all fact-checking articles on archive.org. Fact-checking articles may feature “double-archived” links using both archive.org and archive.is, e.g. “<https://web.archive.org/web/20201229212702/https://archive.md/28fMd>”. Archive.org returns a 404 page for these. To view such a link, please just copy-paste the archive.is part (e.g. “<https://archive.md/28fMd>”) into your browser

6. Answers should be sought from the metadata, any of the sources listed in the fact-checking article (e.g. any hyperlinks to other sites). When that is not possible (e.g. the hyperlinks are inaccessible), annotators should instead retrieve evidence from the internet using the search bar we provide. For image-text claims, fact-checkers retrieve images as evidence. For retrieved images, which are not included in the uploaded images from Phase 1, annotators should ask additional questions for retrieval (i.e., the answers to the question are images). These answers (i.e., images) will be updated to the image candidate list mentioned above. Similarly, if the images are not associated with URL links or the links are not accessible, annotators should use our provided search bar (capable of both image and text search) for retrieval. Answers should be accompanied by a hyperlink to the source, except for the *Image-analysis* and the *Metadata* answering method.
7. Answers can be either *extractive*, e.g. copy-pasted directly from the source, *abstractive*, e.g. written in free-form based on the source, or *boolean*, e.g. written as yes/no with an explanation taken either extractively or abstractively from the source. Where possible, we strongly prefer extractive answers. We also allow *images* to be answers to questions. If an answer cannot be found, we also allow annotators to mark the question as *unanswerable*. We ask annotators to use this instead of deleting unanswerable questions (see Section 4.5).
8. If enough questions have been asked to support a verdict, or if at least ten minutes have passed without the annotator finding enough evidence, a verdict should be given from our for labels described in Section 3.5. Similarly, we ask annotators to provide fine-grained reasons for refuting (as described in Section 3.6) if they find the verdict to be *refuted*.
9. Annotators in Phase 2 should base their verdict on the question-answer pairs they have generated, and *not* on the fact-checking article. Depending on what information has been retrieved, they may disagree with the article.
10. Before proceeding to the next claim, the annotator will be shown a warning with the question-answer pairs they have produced. 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 for the claim.
11. Sometimes, annotators may be in doubt as to whether an additional question should be added to further support the verdict. Generally speaking, we always prefer to have as many question-answer pairs as possible, so if in doubt annotators should err on the side of adding that additional question.

Important! Annotators should not choose a label if the retrieved evidence does not support it; for example, if the label *conflicting evidence* is chosen, there should be evidence documenting the conflict. Labels in Phase 2 can contradict the label of the fact-checker if the annotator believes it is appropriate.

4.1 Claim Correction

In addition to gathering question-answer pairs, Phase 2 also acts as quality control for the claim contextualization and image uploading in Phase 1. This means if Phase 2 annotators encounter a claim that is malformed or not properly contextualized, they can correct it. If Phase 2 annotators find the extracted images of image-text claims inappropriate, they can correct it by uploading a new set of images. The guidelines for claim contextualization are in Section 3.2; the guidelines for image uploading are in Section 3.4; the same criteria hold. Based on our initial review of the data entered in Phase 1, Claim Correction is rarely necessary.

Firstly, we provide some examples of contextualized textual parts of image-text claims that *should* be corrected in Phase 2:

1. The textual claim “Avinash Choudhary claimed Aamir Khan ‘met with terrorist Tarik Jameel and Junaid Shamshed’.”, given the article <https://web.archive.org/web/2021011618>

5501/<https://www.altnews.in/viral-image-aamir-khan-is-not-standing-with-lashkar-e-taiba-terrorists-false-claim/>. The article verifies whether Khan met with the terrorists, rather than verifying the quotation. Avinash Choudhary is the person made the claim to be checked. Here, the Phase 2 annotator should correct the claim to simply “*Aamir Khan met with terrorist Tarik Jameel and Junaid Shamshed.*”

2. The textual part of a claim “*My baby has coronavirus.*”, given the article <https://web.archive.org/web/20201223221522/https://leadstories.com/hoax-alert/2020/05/fact-check-the-little-girl-in-the-viral-photo-asking-for-prayers-does-not-have-coronavirus-the-sign-in-the-picture-is-photoshopped.html>. The extracted textual claim is not comprehensible without the fact-checking article as the pronoun *my* refers to entities mentioned in the article. The Phase 2 annotator should correct the claim to “*The little girl in the image had coronavirus.*”.
3. The textual part of a claim “*Did Sulphur Dioxide Levels Rise In Wuhan Post Coronavirus Outbreak?*”, given the article <https://web.archive.org/web/20210308202718/https://www.boomlive.in/world/did-sulphur-dioxide-levels-rise-in-wuhan-post-coronavirus-outbreak-7026>. The textual part has mistakenly been phrased as a question (probably directly taken from the title of the fact-checking article). The Phase 2 annotator should correct this, following the article: “*Sulphur dioxide levels raised in Wuhan post coronavirus outbreak.*”.
4. The textual part of a claim “*Image showing Barack Obama shaking hands with Hassan Rouhani is real.*”, given the article <https://web.archive.org/web/20200812003618/http://newsmobile.in/articles/2020/01/07/dont-believe-this-fake-picture-of-barack-obama-with-iranian-president/>. The textual part contains a superfluous part of the original claim (e.g., “*is real*”). The Phase 2 annotator should correct this by extracting the claim as close as possible to its original form, rather than the form used in the fact-checking article: “*Image showing Barack Obama shaking hands with Hassan Rouhani.*”. Similarly, Phase 2 annotators should correct the textual claim *No, fishermen drowning is not from Mumbai*, given the article <https://web.archive.org/web/20201108001252/https://www.indiatoday.in/fact-check/story/fact-check-no-video-fishermen-drowning-not-from-mumbai-1569355-2019-07-15>, to “*Fishermen drowning is from Mumbai.*”.

Besides, we also provide some cases that the uploaded images should be revised:

1. The set of uploaded images is not complete. For instance, given the claim in Figure 1, if the previous annotator only uploaded one image, the annotator in Phase Two should correct by uploading both images involved in the claim.
2. The uploaded image is not the original image in the claim. Some may upload the image edited by fact-checkers directly (e.g., the image with a “hoax” stamp as shown in Figure 6). Some may directly upload a screenshot of the claim (e.g., Figure 7(a) is the screenshot of the whole claim while Figure 7(b) is the image part) rather than extracting the images in the claim. For these cases, annotators in Phase 2 should upload the exact images of the image-text claim.

4.2 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 reference to any previously asked question (e.g., there is no coreference in the current question to previous questions).
- Questions should be based on the version of the claim shown in the interface (i.e., the version extracted by phase 1 annotator), 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 necessary questions to gather the evidence needed for the verdict, including world knowledge that might seem obvious but could vary based on one’s background. For instance, Europeans might possess better knowledge of European geography and history than Americans, and vice versa.
- As a guiding principle, at least 2 questions should be asked.
- When asking questions to retrieve information about images, annotators should make the information to be retrieved explicit. For instance, when asking for the publication date of an image, the question *When was the image first published?* is better than the question *What is the origin of the image?*. Similarly, when tracing the publisher of an image, annotators should make the intention explicit by asking a question like *Which website/journal was the image originally published on?*.
- When the question is about the comparison between images, annotators should explicitly mention how the images are compared, by referring to the order of the image, which we will provide (e.g., *Was the photo named 1-th image taken before the 5-th image?*).

The following are examples used to illustrate how questions (the textual part) should be asked. These are based on the image-text claim “*Iran has started to deploy its army, navy and airforce fleets and direct their missile arsenals.*” ²¹:

- Good: What kind of missiles did Iran fire on a US base in 2020?
- Good: What type of missiles are shown in the image?
- Good: Who is the person in this photo?
- Bad: What kind of missiles did Iran fire on a US base? [No time specified to find a statistic]
- Bad: What type of missiles are shown there? [What does *there* refer to? If the question is related to an image, make it explicit]
- Bad: Is it true that Iran has started to deploy its army, navy and airforce fleets and direct their missile arsenals? [Directly paraphrases the claim]
- Bad: What is the origin of the image? [Do not explicitly mention what information is needed when doing media source discovery]

4.3 Question Type

For each asked question, annotators are asked to provide the question type (i.e., which aspect of the claim the question is focused on). Below we provide four options for question types, which are not mutually exclusive as one question can ask for more than one aspect:

²¹<https://web.archive.org/web/20210720124005/https://factcheck.afp.com/photo-iranian-missiles-has-circulated-reports-least-2013>

1. **Text-related:** The question is for checking the text part of an image-text claim and the question is not referring to any image related to the claim. For instance, given the textual part of an image-text claim: *“Israel reserves beach chairs for vaccinated people only.”*, the question *“Did Israel separate beach chairs based on someone’s vaccination status?”* intends to check the textual part independently of the image and thus should be regarded as *text-related*. Another example of *Text-related* questions is a question asking for more details about a textual answer from a previous question. For instance, the answer to a previous question is that *“The image shows Bonifacio Global City.”* Then the next question *“Which country is Bonifacio Global City in?”* is regarded as *Text-related*.
2. **Image-related:** Image-related questions are commonly seen during the fact-checking process for image-text claims. These questions can be asking about various kinds of information of images (e.g., the date of publishing, the location for taking the image, entities in the image and events in the image). Sometimes, tracing the publication source of images can also be used for refuting an image-text claim. For instance, given an image-text claim, the answer to the question *“Which platform was the image originally published on?”* is *Artsi*, a website for selling AI-generated artworks. An image-related question can be used to refuting a claim if the latter considers the image to be real. As long as a question is asking about information of images, it should fall in the category of *Image-related* questions. The images need to be in the image candidate list which contains images from the claim and images retrieved as answers to previous questions (More details in Section 4.6). Annotators are asked to select images from the candidate list when asking *Image-related* questions.
3. **Metadata-related:** Questions about metadata can be used to highlight particular aspects of the claim, in order to reason about the publication date or publication source. For example, if an image-text claim *“Photos show recent farmers protest in Pipli, Haryana.”* with photos taken in 2014, was published on September 10, 2020, the date of publication can be used as evidence for refuting. In this case, annotators can generate a question *“When was the claim made?”*, which is regarded as a *Metadata-related* question. Similarly, questions about publication source can be used to refute satirical claims. – *“Where was this claim published?”*, *“www.theonion.com”*. Note, when asking for metadata of an image in the claim (e.g., *When was the photo taken?* or *Who published the image?*) the question is *Image-related*. Only when questioning about the metadata of the overall claim (e.g., *when was the claim made?*), it can be described as *Metadata-related*.
4. **Commonsense-related:** As a part of the question-generation process, annotators may have to make assumptions and/or use world knowledge to interpret the claim. Due to differences in background, annotators should decide if some facts are beyond an average English speaker is likely to know (e.g., *“Canada is the third-largest country in terms of land mass”*), and should consider asking questions about these facts. In some cases, commonsense involves a politically charged judgement. For instance, to verify the claim *“Kamala Harris is a black American.”*, it calls for the definition of *Black American*. Some may define *Black American* as persons “having origins in any of the Black racial groups of Africa... and Afro-Caribbean entries, such as Haitian and Jamaican”, while some may think black people in America are Black American. 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 *Image-related* questions, annotators should also specify which images are involved in the question. We provide an image list for annotators to select from, which contains images of the image-text claim as well as images which are answers to previous questions. We prohibit annotators from asking questions about images not in the image list (i.e., questions should not be about images not shown in the original claim nor answers to previous questions).

In some cases, fact-checker conduct media source discovery to find a cleaner version of the image in the claim first and ask questions based on the cleaner version. In such cases, annotators should convert this

process into a question (e.g., *What is a cleaner version of the image?*) and ask a separate question to retrieving the cleaner image. The expected answer, which may be a similar image to the original image with higher resolution, will be added to the image list so that further questions can be asked about the cleaner version of the image.

4.4 Answering Method

To find evidence for each question, annotators may use different answering methods. To answer a text-related without any images, textual search is sufficient. However, when answering some image-related questions, they may retrieve information of images via image search, the inputs to which are images. Image analysis (e.g., comparison between images or checking details of a single image) could also be involved when answering image-related questions. Metadata would also be used for answering metadata-related questions.

As we mentioned at the beginning of the section, answers (in the format of either texts or images) can be sought from the source listed on the fact-checking article, as long as it is still accessible. Regarding these cases, there is no need for annotators to conduct search themselves. However, for these questions, we still ask annotators to describe how the answer could be found, based on their own judgement. Below we provide some examples how different answering methods could be applied to questions:

1. **Text Search:** Text search is the case when the inputs to the search engine would be text only. When answering questions related only to the textual part of image-text claims, text search is commonly used. However, text search could also be used for answering image-related questions. For example, when tracing the source of an image, fact-checkers occasionally use keyword search to find the source.
2. **Image Search:** Image search is the case that the inputs to search engines would be images, e.g., reverse image search, and is a common method for retrieving information about images. If no images would be involved during searching, annotators should select *Text Search*.
3. **Image Analysis:** Besides retrieving open-world information about images, image analysis is often involved for answering image-related questions. In some cases, the question is about some details of a single image (e.g., the race of the person in the image). Meanwhile, image analysis is frequently involved for comparing two images (e.g., whether the landmark in an image is the same as the one in Google maps; whether the persons in two images are the same).
4. **Metadata:** When answering metadata-related questions, annotators should choose this category as the answering method.
5. **Unknown:** If annotators find it hard to justify the answering method, they can choose this category as the answering method and provide explanations.

For both *Text Search* and *Image Search* answering methods, annotators should provide the URL of the searching result.

4.5 Answer Generation

To find answers to the questions, the annotators can rely on metadata, or any sources (e.g., images and articles) linked from the fact-checking site. Where these do not contain appropriate information, either because they are not relevant to a 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 it (and only when they are not from fact-checking websites). Similarly, other fact-checking articles found through searches should be avoided.

We provide two kinds of search functionality. One is responsible for text search while the other for image search (i.e., takes images as inputs). Annotators should choose different search functionalities according to the format of the input.

When using the image search functionality, annotators can only use images from our provided image list as inputs. The image list is initialized with images in the image-text claim (i.e., uploaded images by the annotator in Phase 1). When an answer to a question is an image, the image (i.e., the answer) will be automatically added to the image list. If the image they would like to use is not in the list, they should ask an additional question to retrieve the image first.

4.6 Answer Type

Once an answer is 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. Extractive answers could also be sourced from images. For instance, given an image and a question "*what is the text on the flag in the image?*", the extractive answer is the extracted text of the flag.
- **Abstractive:** A free-form answer can be constructed based on the source, but not directly copy-pasted. Similarly, abstractive answers could be sourced from images (e.g., the question is "*what is the color of the building?*" and the answer is "*It is red.*").
- **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..."). Boolean answers could be sourced from images as well.
- **Image:** The answer is an image. For instance, fact-checkers usually conduct reverse image search for finding a cleaner version of the original image.
- **Unanswerable:** No source can be found to answer the question.

If the source is not an image, extractive answers are preferred to abstractive and boolean answers. For image answers, annotators should upload the image instead of typing textual answers.

In some cases, annotators might find different answers from different sources (e.g., annotators use both reverse image search and text search with keywords for retrieving different sets of relevant images). Our annotation tool 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 it. In such cases, please provide the partial answer rather than marking the question as unanswerable.

Our search engine indicates 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. If an annotator wishes to use information from such a source, we strongly prefer that they find similar, corroborating information from an additional source 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 the questions and answers.
- DO NOT use the fact-checking article itself, or any other version of it you find on the internet, as evidence to support an answer.
- DO NOT submit answers using other articles from fact-checking websites, such as politifact.com or factcheck.org, as evidence.

- DO NOT simply refer to the source as an authority in abstractive answers (and boolean explanations), e.g. do not use answers like “no, because the post on Flickr contradicts.”. Rather, write out what the source says, e.g. “no, because the original post of the image shows the image is taken in Mumbai.”. If you consider it important to mention the source, write that the source says – e.g., “no, because the original post of the image on Flickr shows the image is taken in Mumbai.”.

4.7 Reasoning Chains of Claims

Annotators can build up reasoning chains across multiple questions, meaning that answers to one question can be used in the next question. Frequently, image answers are used as inputs to subsequent questions. For instance, the first question is asking for retrieving a cleaner version of the original image and the retrieved image answer will be used as inputs for subsequent questions. For another instance, for the claim “The photo shows Philippines.”, the first question is “Which city does the image show?”. The answer is “It shows Bonifacio Global City.”. Based on the answer, we can further ask the second question to get more concrete evidence, “Which country is Bonifacio Global City in?”. Note that while the *generation* of the second question assumes knowledge of the answer to the first, it is *understandable* without it.

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). Therefore, if they disagree with the fact-checker about the label, they can select a different label.

5 Phase 3: Quality Control and Justification Generation

Given a claim with associated evidence, we ask a third round of annotators to give a verdict for the claim. Once we have collected evidence in the form of generated questions and retrieved answers, in Phase 3, we want to provide a measure of quality. Given an image-text claim with associated evidence, we ask a third round of annotators to give a verdict for the claim. Besides, we would like the annotators to write a justification (i.e., a short statement justifying their verdict) for each claim. Crucially, the annotators at this round do not have access to the original fact-checking article, or to the claim label. Below is an overview of the phase and details are provided in the following sections. Further documentation can also be found on-the-fly using the tool-tips in the annotation interface.

1. Annotators should first read the claim (both the image and the text part), the metadata, and the question-answer pairs (answers to questions could be in the format of images and image-related questions are paired with images). This is the only information which should be used during this phase.
2. It is important that annotators in the quality control phase do not use web search to find additional information or rely on background knowledge which an average English speaker might not have. Commonsense facts that are known to (almost) everyone can be used – see Section 5.1.1.
3. If the claim or any of the question-answer pairs lack context, they can be flagged. This helps us diagnose what went wrong with a set of question-answer pairs in the case annotators disagree over the label.
4. After reviewing the claim and the question-answer pairs, annotators should assign a label to the claim (see the four labels introduced in Section 3.5).
5. Finally, annotators should write a short statement justifying the verdict. Considering images may also be used as part of justifications, we provide unique index for each image (the images of the claim and the images in answers) and annotators can refer to indexes of images (e.g., *In [CLAIM_IMG_0], the person is ...*). If any commonsense information (e.g. background knowledge

which an average English speaker *is* likely to have) is used to give the verdict, but that information is not mentioned in any question-answer pair, it should be mentioned in the justification. For advice regarding justification production, see Section 5.2.

5.1 Quality Control

Given an image-text claim associated with its evidence in the format of question-answering (QA) pairs, the annotators should first check the QA pairs. Regarding the question, the annotators should check whether it is readable and relevant to the claim. For image-related questions, annotators will need to check if the images are really related to the question. For the answers, the annotators should also check if there are potential problems with the answers (by selecting options provided on the annotation platform).

5.1.1 Commonsense Knowledge

When giving a verdict, annotators sometimes need to rely on commonsense knowledge. Here, we consider only basic facts that an average English speaker is likely to know – e.g., “*Earth is a planet*” or “*raindrops consist of water*”. No other information beyond the question-answer pairs can be used in this phase.

We ask annotators to be relatively strict with what they consider commonsense, but use their own judgment. For example, we would consider “*Canada is a country*” commonsense, but not “*Canada is the third-largest country in terms of land mass*”. If an annotator is in doubt as to whether something is considered commonsense, they should not consider it commonsense.

5.2 Justification Generation

In addition to the verdict, we also ask annotators in Phase 3 to write a short statement justifying their verdict. This justification should explain the reasoning process used to reach the verdict, along with any commonsense knowledge. If calculations or comparisons were used, e.g. “*6.3% is greater than 6.1%*” or “*10-4=6*”, they should be explicitly stated in the justification. Similarly, any rounding logic – e.g. “*4.3 million is approximately 4 million*” – should be explicitly stated here.

Other than commonsense knowledge, there should not be any new information presented in this statement. The justification should only describe how the annotators used the information present in the claim, the metadata, and the question-answer pairs to reach their verdict. If a verdict cannot be reached, e.g. if the *not enough information* label is chosen, annotators should instead describe what information is missing – e.g. “*I cannot determine if Canada is the third-largest country, because the questions do not specify how large other countries are.*”

Similarly, in cases of conflicting evidence, annotators should describe which questions and answers lead to the conflict, and how they contradict – e.g. “*This claim is cherry-picked as it looks only at the price of vanilla ice cream, for which an increase did take place, but leaves out other flavors, where no increase happened.*”

Considering the fact that justifications for image-text claims may include images, we provide unique index for each image uploaded by annotators in Phase Two (both the images of the claim and the images in answers). If annotators think it is necessary to include images as part of the justification, they can refer to the index of the image with a bracket. For instance, if they want to use an image indexed as *ANS_IMG_0*, they can refer to the image with its index and a bracket as: “*In the [ANS_IMG_0], the person is ...*”.

6 Dispute Resolution

For some claims, there may be a disagreement between the labels produced by the annotators in the question generation and quality control phases. In those cases, the claim will go through a second round

of question generation and quality control. While the instructions given in Sections 4 and 5 still apply, we give a few extra recommendations specific to dispute resolution here.

6.1 Vague Claims

Some claims may pass to the dispute resolution phase because they were too vague for annotators in phases 2 and 3 to agree on the meaning. In order to catch these cases, the final step of dispute resolution – that is, the extra quality control step at the end – includes an additional label, *Claim Too Vague*. This should be selected if and only if an annotator can understand the claim (i.e. it is readable), but it is open to different interpretations. For example, the claim “*Ohio is the best state*” is too vague as it is not clear what “best” refers to.

6.2 Adding and Modifying Questions

The aim of dispute resolution is to resolve the conflict so that a potential new reader can come to a conclusive verdict. As such, the annotator should not necessarily agree with either the Phase 2 or the Phase 3 verdict; they should attempt to make the fact-checking unambiguous. There may be cases where new questions need to be added, and cases where existing questions should be changed but no new questions are necessary. There may also be cases where no change to the evidence is necessary at all, but where either the Phase 2 or Phase 3 annotator has simply entered a wrong verdict. For this final category adding additional evidence to provide clarity can still be helpful, but it is not necessary; annotators should use their own judgment here.

6.3 NEE-verdicts

A common case for dispute resolution is the situation where the Phase 2 annotator has selected *Supported*, *Refuted*, or *Conflicting Evidence/Cherrypicking* as the verdict, but the Phase 3 annotator has selected *Not Enough Evidence*. This can happen for example if Phase Two 2 forget to gather some of the evidence they use to reach the verdict, rely on aspects only stated in the fact-checking article without making it explicit through a question-answer pair, or overestimate the strength of the evidence they have gathered. In these cases, the aim of dispute resolution is to gather additional evidence and resolve the conflict that way; i.e. it is not sufficient to give a *Not Enough Evidence*-verdict without attempting to add evidence (although the same time limit as in Phase 2 applies).

References

Grace Luo, Trevor Darrell, and Anna Rohrbach. Newsclippings: Automatic generation of out-of-context multimodal media. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP*, pages 6801–6817. Association for Computational Linguistics, 2021.