

Uolo Base Guidelines V.5 and V.6 Basic Labeling

Contents

Purpose of Uolo	3
Big points to keep in mind:	3
Immediate Escalation:.....	4
How to review content:	4
The Basics of Rating.....	7
What is a central claim?	8
What is a statement of fact?.....	8
Is there a central claim?	8
A note about ads	10
Other circumstances where there may not be a central claim (for project purposes).....	11
What is the central claim?	12
Multipart Claims.....	12
Combining related claims.....	12
Most severe claim	13
Emphasized claim.....	14
How to search for and find appropriate evidence	15
How to determine the final evaluation	19
Putting It All Together	21
Impact Questions	21
The Satire and Opinion Labels	25
Satire.....	25
We ask you to tell us whether the content you are evaluating is Satire:	25
The Likely to mislead and not likely to mislead options for satire posts	25
Opinion	29
Claim Location Label	31
More on Evidence Labels	32
Evidence Dos and Don'ts.....	32
Inappropriate Evidence.....	33
Trustworthiness Labels	34

More on Final Evaluations	35
Tips and How to Avoid Pitfalls	37
Writing claims like a pro	37
Photo or video is the claim	38

Confidential
129889264

Purpose of Uolo

To develop a system that identifies and matches evidence to Misinformation.

The process is split into 2 tasks, Task A and Task B. Task A is identifying if the job has a claim, Task B is where you state the claim, find evidence relating to the claim and evaluate the truthfulness of the claim.

Task A

The first step of the rating process where you make key decisions about the content such as determining the main purpose of the post, is it satire, and should it be passed on to Task B for a full evaluation.

Task B

The second step of the rating process where you identify and select or write the claim, find evidence, assess the quality of the evidence, and make the final evaluation.

Big points to keep in mind:

This document contains ever-evolving and ever-improving guidance on how to think about the task's questions and what the response options mean for researching a piece of **content**. One piece of content may contain a single or in some queues, many media components such as text, photo, video, and links. To refer to any one component in a piece of content (e.g., one video, or one photo), this guideline uses the term **"content component"** or more simply "component".

In this document, the separate steps of evaluating a job are broken down, but you should always keep in mind the main purpose of the project and use your best judgment on every content. What is the content really trying to say, and is that main point a factual claim that is definable and can be evaluated?

- **This isn't intended to be a strict set of guidelines, because this isn't intended to be a purely objective task.** There will be some variability between people in how they approach the task and how they respond to the questions, and that is by design. However, try as much as possible to follow the instructions laid out below and to answer each question to the best of your ability.

Immediate Escalation:

If the content contains imagery or text/voice indicating or soliciting Child Exploitation or Child Nudity, escalate the Job ID immediately to your manager.

- Child Exploitive Imagery (CEI) refers to imagery (images, videos) depicting the sexual exploitation of a child.
- If you are not comfortable reviewing the content, you can skip the job after escalating. ▪ **IMPORTANT:** CEI should never be screenshot or replicated in any way as it only further exacerbates the issue. Please always look to use task/job number for issue identification.

If you are not comfortable reviewing the content on the webpage for any other reason, please escalate the job ID to your manager and skip to the next job.

How to review content:

Some queues have only jobs with a single content type (plus the text from post author, but others may contain multiple content types.

When reviewing jobs with multiple content types, it is essential to review all the text, image, video, and link components for each content. Different contents will be made up of different combinations of content components.

- **For links:** Read through the title and first few lines of text (or watch a bit of the video if one is included in the link). If there is a main photo or video on the linked article, if the authenticity of the media is related to the main point of your central claim, this must also be included in the claim and be a part of your evaluation. For some markets, this is a very common content type that often hides misinformation. The post will contain a link and the article title, the linked article then contains a false image that does not relate to what is being described in the article.
- **For videos:** Look over any text caption that accompanies the video and watch through the first 30 seconds of video content. You may need to watch for a bit longer or need to skip to watch a few seconds of other portions in the middle or end of the video in some cases. It may also be helpful to reference the video transcript, when available. You do not need to watch videos in their entirety.
- **For images:** Consider both what the image is portraying as well as any overlaid text. Always consider any post text by the author as it relates to the image as well. Sometimes the image is the claim. A rule of thumb to use to determine if the image itself is a claim is to describe what statement the image is conveying and if it is something you would evaluate if it were written in text, you would

evaluate the same claim, as it is implied by the image.

- **For text from post author on all content types:** Read the text and review any accompanying media components as outlined below. **When hashtags and tagged mentions are part of the post text, please consider them as part of your evaluation.**

- o Some contents will use **irrelevant or overuse hashtags** or tagged mentions in a spammy manner. When this is the case, please do not consider them in your evaluation.

The central claim is often described in one or more of these locations in a post:

- **Text component:**
 - The caption on the content (often referenced as text from post author) typically accompanying an image, link, or video.
- **Image component:**
 - The image or overlaid text on the image.
- **Link or article component:**
 - In the headline or title, or body text of the landing page article.
- **Video component:**
 - The video, overlaid text, or audio

Example:

This content contains **three** components:



Anatomy of a Post

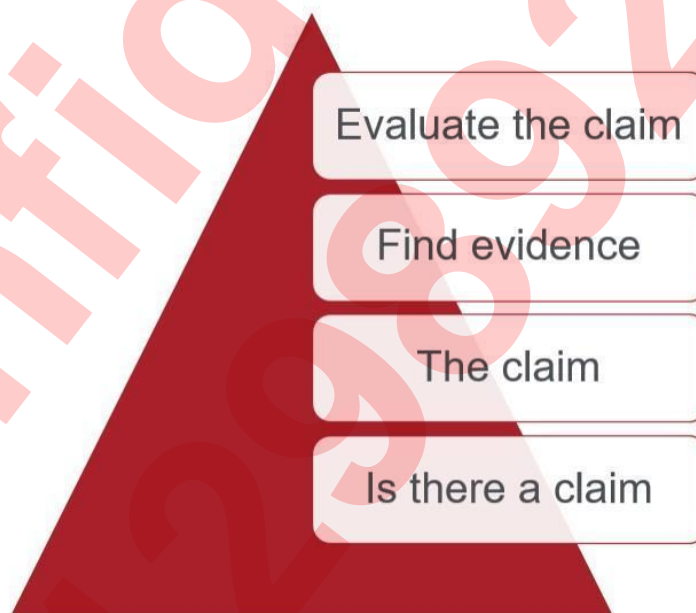
- Once you are done reviewing all components of the content, whether just one or many, you should consider whether there is a central claim across the total set of components, holistically.

The Basics of Rating

Key concepts to understand for successful rating on the project.

There are many rules and many labels on project Uolo, but at the core of the project, there are four main areas to understand. We recommend that as you start your journey learning the project and reading its guidelines, you keep in mind everything about the project and all you will be doing, ties back into the basics. Like a pyramid, one step leads to the next and to the primary goal of an appropriate final evaluation. We are like detectives looking to determine the veracity of social media posts in a methodical way. If early steps in the process are inappropriately completed or missed, the primary objective of evaluating a claim as true or false (or somewhere in between) will not be met.

Rating all starts with the claim. For example, you will not catch a false claim if you said there was no claim in the first place, or if you did not identify the correct claim or find evidence for it. With that in mind, these are the key concepts of rating that build upon one another.



What is a central claim?

The central claim is a statement of fact relating to the content's main point or purpose that can be supported or contradicted (or somewhere in between) using an evidence source.

Some jobs will only have one statement of fact being made, while others might have multiple.

What is a statement of fact?

A statement of fact is a sentence or main idea that the content asserts as a fact (regardless of whether the statement is factually accurate).

Is there a central claim?

Determining if there is a central claim in a post starts with understanding the main purpose for the post. You will hear a lot about the main purpose on this project and you should know the term is used in two different contexts, one you will learn more about later in the "what is the central claim" section, but for purposes of deciding if there is a claim, the main purpose of the post refers to a broader level.

In the context of determining if there is a claim, we are speaking of the main purpose in a broad and overall sense, referring to a broad category of posting:

- ◆ is the main purpose to inform the reader with a fact or facts?
- ◆ is it to persuade the reader (to agree with an opinion, purchase a product, sign a petition, etc.)?
- ◆ is it to entertain the reader (a joke, a well wishes post, etc.)?

While we sometimes find a central claim based on a statement of fact in entertainment/persuasive content, there must also be an element of "to inform". Thus, if the content's main purpose is to persuade us to agree with an opinion, it often also informs us of something to prompt that persuasion. Same concept with content meant to entertain: does it inform us of something that can be researched and shown to be supported or not supported to set up the entertainment?

****Content that has the purpose of informing us is suitable for our evaluation. The caveat for content meant to persuade or entertain is that it should also be informing us of something related to its main purpose.**

The post has a central claim when:

- The main purpose of the post is strictly to inform the reader of a fact or facts that can be evaluated.
- The content is opinion/persuasion based on facts.
 - **Example:** Biden is the worst president ever because he has caused gas prices to soar. The opinion here is based on a statement about the cause of rising gas prices that can be evaluated.
 - **Example:** An ad that states: Our green tea supplements are the most effective. Studies show that taking green tea supplements daily reduces blood sugar by 50%. The persuasion is based on a claim that can be objectively assessed outside of product information.
- The entertainment is based on facts.
 - **Example:** A photo of a St. Bernard's tongue with the caption "adorable floppy tongue breaks the world record for the longest". While the main purpose of the photo post is to entertain, the entertainment component of the post is premised on claiming a fact that can be checked.

When might the content NOT have a central claim:

- The content does not contain a central statement of fact that relates to the **MAIN PURPOSE** of the post that can be supported or contradicted (or somewhere in between) by evidence.
- This may be because:
 - The main purpose of the content is solely to convey an opinion or values statement and it does not contain any statements of facts as rationale/support for the opinion/value statement; or
 - The main purpose of the content is not related to conveying a statement of fact.
 - **An entertainment example:**
 - The post is an entertaining video of a dog playing fetch during the 2020 quarantine for Covid and the caption is "he sure misses when I was home playing fetch all day during the Covid quarantine." – In this case, the main purpose of the post is to entertain you, not inform you there was a Covid quarantine in 2020. It would be marked as having no central claim.
 - **A persuasion/opinion example:**
 - The country has been going downhill for decades now. Every

election, each administration gets worse and worse. I am convinced we have not had a good president since Teddy Roosevelt in the early years of the 20th century. - Here, the main purpose is not to inform that Roosevelt was the president in the early 1900s. This post would also be labeled as no central claim.

- **A religious post example:**

- Religious books such as the Bible and Quran are typically subject to several different interpretations, so any posts that make claims such as the Book of Job says tattoos are a sin are labeled as no claim, as there is no definitive way to fact check such a claim. Thus, sermons, prayers, and any other religious or spiritual interpretation posts (including astrology) are typically labeled as no claim. Exceptions include content with the main purpose of informing the reader of a verbatim verse such as "Psalm 23 says..."

A note about ads

Ad content is not automatically labeled as having no central claim. Firstly, all components of the content you are labelling must be considered for part of the potential claim. So, for example, if there is a photo showing on SRT with some post text and a link, you must also navigate to the external link in addition to considering what is showing on SRT itself.

To determine if an ad has a claim for project purposes, please consider the following. It is out of the scope of the project to evaluate pricing or effectiveness of a product or service, but if there is a claim that can be evaluated outside of product specific information, it can and should be evaluated. For example, is a post an ad for a lemongrass supplement that is said to promote weight loss? If so, it can be evaluated based on its ingredients. We can evaluate the claim that lemongrass (not the product) promotes weight loss.

External Link Name

👉 Grab Your Keto-Gummies From Oprah Until the 30th!

External Link Caption

Learn more

External Link URL

<https://www.millertopketo.us.com/dsdss>

****Remember:** You should consider all components of a given content (i.e., any post text, images, video, audio, overlaid text, headlines/titles of articles/links, and body text of articles/links) to identify the central claim (or if there is one) of the content.

Other circumstances where there may not be a central claim (for project purposes)

If you cannot determine if there is a claim:

There are three primary reasons you may not be able to determine if there is a claim, usually because of access issues or language barriers.

- **SRT Access issues**
 - You cannot review the content due to a bug or technical issue in SRT. This may be because the SRT preview is broken and/or the content or caption does not render properly.
- **Content inaccessible**
 - You cannot review the content because it was not possible to access content that is NECESSARY for identifying the claim. This may be because:
 - the link is broken, the content is missing, or you encounter a 'page does not exist' message, your web browser blocks the link or identifies it as unsafe, or the website asks you to pay for content.
- **Wrong language or lacking cultural context.**
 - You cannot review the content because it is in a different language from what you are assigned to rate.
 - Please note that automatic translations that appear on SRT do not count as part of the content you are evaluating and should not be considered for claim assessment.

What is the central claim?

When it comes to forming a central claim for content where you have identified the main purpose as at least a partial intent to inform the reader, the base consideration is where our second definition of the main purpose come into play.

What is the poster **REALLY** trying to say? If you consider the whole post in context, what's the main point? Let's say the post is about the economy and the post text states that inflation is rising at an alarming rate and goes on to say that the price of bread has increased, gas prices have increased, and peanut butter went up in price a dollar since this time last year. The main purpose of that post is not simply to say that peanut butter prices are rising, and that should not be your claim. In that case, your claim would be something about inflation rising in general with the prices of gas and groceries being major indicators.

After you have evaluated **ALL** the components of a job and have determined that the main purpose of the post is to **inform** or at least partly to inform, decide which claim, or claims are related to that main purpose. What is the author trying to say with their post? What point are they trying to make?

Multipart Claims

Important things to remember:

Combining related claims

- Sometimes, a piece of content will have one primary central claim or a single multi-part claim, where the parts are closely related.
- For example, if a post says: "Lactose intolerance is more common than any other allergy worldwide. It's estimated that up to two billion people experience symptoms related to lactose intolerance annually." You may identify these claims:
 - Claim 1: Lactose intolerance is more common than any other allergy worldwide.
 - Claim 2: It's estimated that up to two billion people experience symptoms related to lactose intolerance annually.
- In this case, the second claim builds upon the first claim, so you would **synthesize a unified central claim**: "Lactose intolerance is the most common allergy in the world and up to two billion people experience related symptoms each year."

Most severe claim

In some cases, however, a piece of content may contain multiple distinct central claims. In these cases, we ask you to focus on identifying and evaluating **the most severe claim**.

One of the most common scenarios we see as far as necessitating selecting the most severe claim are the very typical posts where we must choose between what was said or did they say it. Consider the following example:

- Social media often posts linked media or content that frames a claim as “Minister X stated that (insert claim)”. These post types can be tricky because at their core, they are multiple claim contents where you must decide if the claim is that Minister X said this as reported or if what Minister X said was true. You also must consider the main purpose of the post.

In example 1, the most impactful/severe claim is the truth of what was said, and it also relates to the main purpose of the post.

Sinan Ogan announcing elections will be postponed.



In the above example, the main purpose of the post is to inform the reader that elections are postponed.

Appropriate approach to the claim: “Turkish 2023 elections are postponed”.

One tip for making this determination includes revisiting some of the methods of identifying a claim when there are multiple claims. In this example, thinking in terms of which is the most 'severe' claim is appropriate. Between whether he said that or if elections are really postponed, election postponement is the most impactful and severe claim.

In example 2, the most impactful/severe claim is if Putin made this statement as quoted (what he said is also not falsifiable) and it also related to the main purpose of the post.



In the above example, the main purpose of the post is to inform the reader Putin made the statement shown in the image. Considerations here are that the statement Putin made is not something that can be falsified, as the creation of Ukraine is an abstract point that would have several different answers from scholars. However, the fact that Putin is an impactful world leader, we also need to consider that impactful people make impactful statements. If the main purpose of the post emphasizes informing the reader that *THIS* comment was made, quote verification should be included in the claim.

Appropriate approach to the claim: ***"Putin stated that Ukraine is not a neighboring country, and that it was created by Bolshevik communist Russia"***.

Emphasized Claim

- If all the claims are equally impactful and relevant to the content's main purpose, you should evaluate the claim that is most apparent or obvious as the central claim (for example, a claim in a bold font, or a claim that is circled in red).

How to search for and find appropriate evidence

General tips for finding evidence

Most importantly, your central claim matters. Finding evidence for all or most of the jobs you rate all starts with a properly formed and accurately reflective central claim. If you form your claim improperly, for example, “This is news about Covid in New York” and you search for “Covid news New York” you are not likely to yield results that are specific enough to directly address the real claim you have in mind.

Step one is always form your claim to be specific and in your own words

- **Do not** use broad terms like “news about”, “Covid vaccination”, “Biden said” or “Putin lied”.
- **Do not** copy and paste the post text by author to use as your central claim.
- **Use a specific central claim** that directly addresses the main thing you will be researching, which is the main thing the post is about.

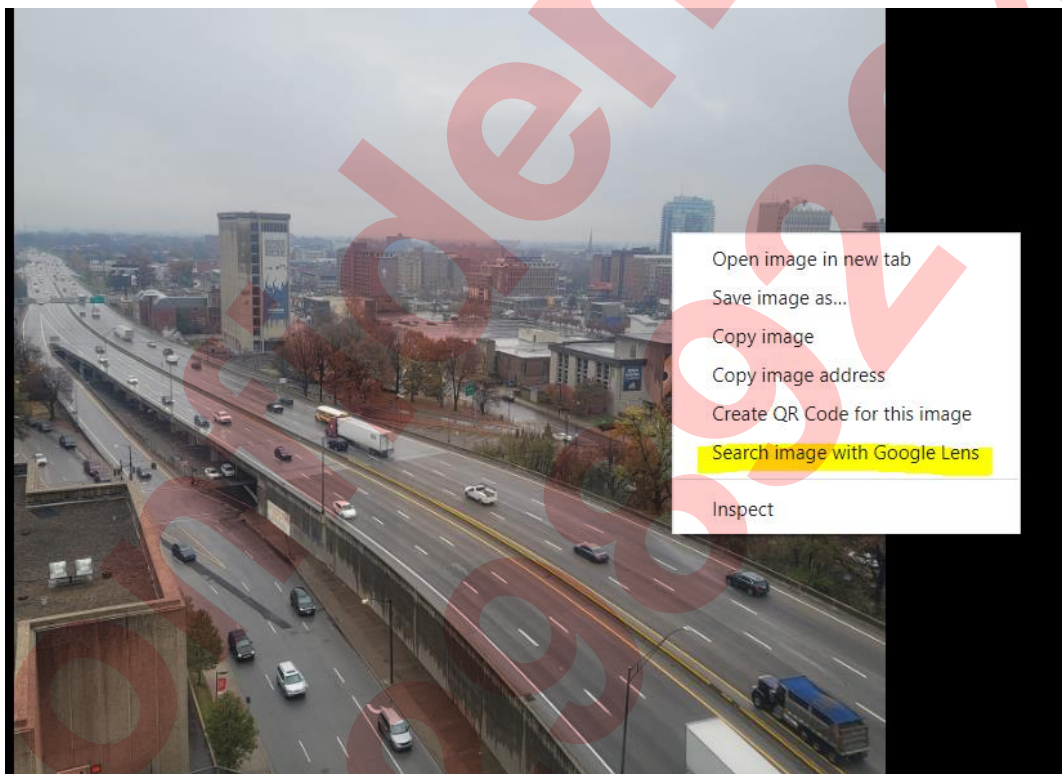
Reverse image search is often very useful if your claim contains an image component

You can do that by following these steps:

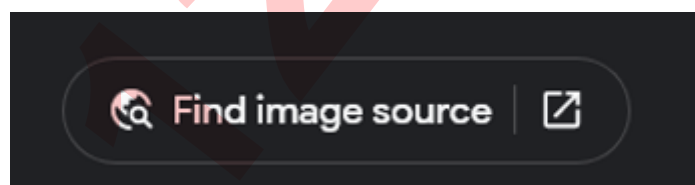
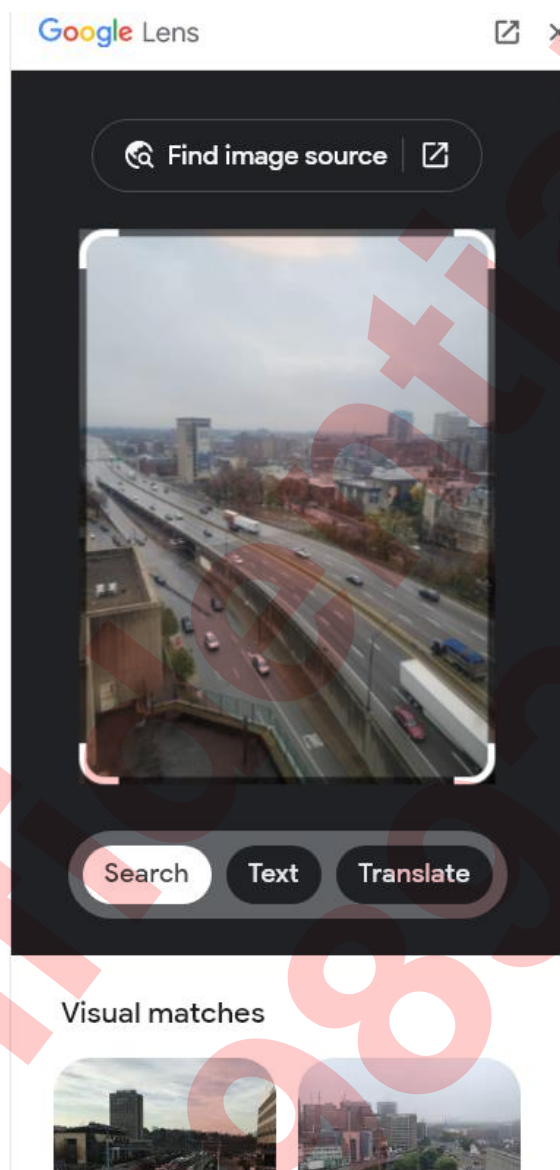
With so many photo claims, you will often find the best way to find evidence is to do a reverse image search for any photos on the post.

You can do that by following these steps:

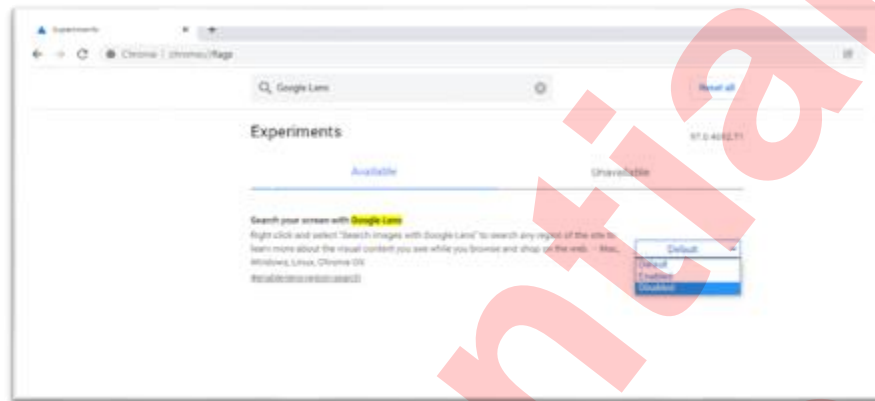
1. Click on the image contained within the post to enlarge it so it fits the full screen, and then right click with your mouse and the following box appears.



2. Select "Search image with Google Lens". It will bring up a Google Search box result like the one below.



3. Select "Find image source". It will open a new window for a Google search results page.

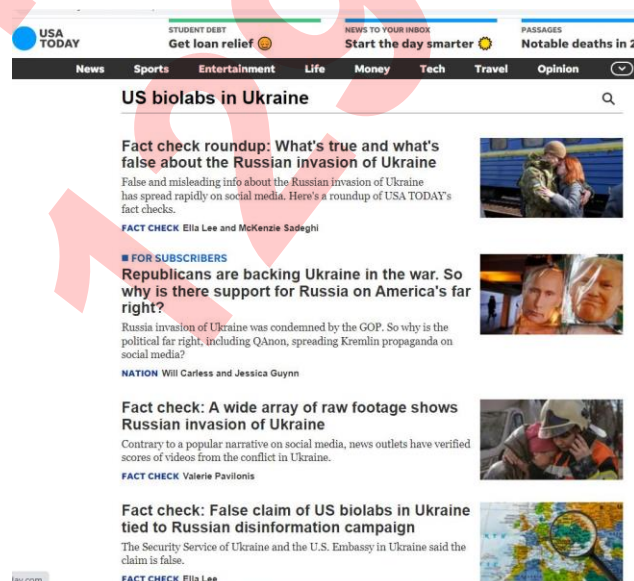


You can also add keywords to an image search to increase your chances of finding evidence.

Another method to try is to go directly to an appropriate news site, for example, maybe you have had good experiences with finding evidence on Reuters, USA Today, or AFP (or any other site of your choosing) and search for your keywords directly in the search bar there.

Search directly on a favorite site

For example, say you are looking for evidence for a specific claim about US Biolabs in Ukraine. Perhaps you have had good results recently from USA Today (or any other site of your choosing), you can try searching for US Biolabs in Ukraine directly on USA Today as shown below:



How to determine the final evaluation

Once you've formed and written your central claim and found credible evidence, it's time to give your final evaluation of the job.

Please take care to familiarize yourself with the wording of some of the final evaluation labels and assure you understand how to differentiate fully, partly, and not contradicted for content that is labeled as Partly supported or Not Supported.

- **Fully supported** – Based on the evidence you found; the central claim is fully supported. *(This means the major components of the claim you found is 100% true, no part of the central claim is incorrect or false.)*
- **Partly supported** - Based on the evidence you found, the central claim is “partly supported” (*partly true*), but also:
 - **Partly contradicted**: Part of the central claim is true, but part of it is proven to be false.
 - **Not contradicted**: Part of the central claim is true, but there is absolutely no evidence available for another major component of the claim implied or otherwise.

Tips for using partly supported/not contradicted

Keep in mind that to use this label a major part of the claim must be supported, and it must be a multi-part claim where there is no evidence for a second and equally severe component of the claim. If there is a severe claim that is supported, then you should consider using the fully supported label instead.

However, let's consider the hypothetical content that says global inflation has risen by the highest rate since 1972 but the World Economic Forum predicts that it will decrease throughout 2023. You might rightfully consider each of those components of the claim equally impactful. Now, let's say you can find evidence that confirms the World Economic Forum made that statement but regardless of how hard you look, you can find nothing that confirms or refutes the statement that inflation rose at the highest rate since 1972. It is appropriate in this scenario to consider the partly supported/not contradicted label as a major part of the claim is supported and there is no evidence available for the other equally impactful portion of the claim.

- **Not supported** - Based on the evidence you found, the central claim is not supported (*not proven to be true*) and is also:
 - **Fully contradicted**: The entire central claim is completely proven to be false/incorrect/untrue.
 - **Partly contradicted**: Only a small part of the claim was proven to be true or there was no evidence, but most of the main component of the claim were shown to be false, contradicted, or highly misleading.

Tips for using Not supported/partly contradicted

Conversely from the partly supported/not contradicted label, the not supported and partly contradicted label is highly useful when a major component of the content is not supported and is some degree of contradicted and there is another equally severe component that has no evidence.

Please consider this hypothetical example: Content that claims that there is a recall on a dog food brand in France and Spain because several dogs have died after eating it. If you can find evidence that this did not happen in France and that component of the claim is not supported but you also consider this happening in Spain as part of your claim and there is no evidence to be found, you might consider the not supported and partly contradicted label as appropriate.

- **Not contradicted:** see note below.

***Note on the “Not Supported/Not Contradicted”:**

Infrequently, there are articles, fact check or otherwise, that do address a widely reported claim or topic to explain that it is truly not known whether it is true or false. Generally, it is when news outlets want to make it a point that the question is open to discussion or interpretation because the facts are unknown, and that people should be careful to assume it is true, but also not too quick to assume it is false. Another way to think of this is that to use the label, the evidence should clearly state that the claim remains an **unknown**. This is what we are looking for when using “not contradicted”.

“Not Supported, Not Contradicted” is meant to be used in a rare case when there is an evidence article that specifically addresses the claim but specifically makes it a point that we cannot or should not say whether it is true or false.

****It is important to note that is an evidence source simply states there is no evidence of a claim, it does not automatically warrant a not contradicted label. Often, fact checking articles use this language to state say the claim is not true. Overall context of your evidence statement should be considered. Further, if you can find no evidence about your identified claim, the job should be marked as such and not labeled with a not contradicted label.**

Putting It All Together

Now that you know the basics of rating, we will address how to approach other labels you may encounter depending on which queue you are working through. Please note that the questions you see on the labeling UI vary and you may not see every question on every job.

Impact Questions

Some queues, particularly in Task A, you would see questions about the claim might impact someone's life if it turned out to be wrong.

Select all that apply. If this claim turns out to be wrong, it will impact someone's:(Multiselect)

- a) **Health, safety, or well-being.**
- b) **Political beliefs or behavior**
- c) **Views of other groups, such as race, ideology, or religion**
- d) **Views of the government, the news media, or other institutions**
- e) **Something else:**
 - I. Please describe in a few words:

PURPOSE: The purpose of the above question is to determine what kind of an impact it would have on people's lives if it turned out to be wrong.

Guidance for the question:

- Your answer to this question should represent the whole content you are evaluating. To answer the question, you should consider all components of the content you are evaluating, including any text captions in or on media content (link/photo/video), any video audio or transcript, and any post text.
 - For content with a video component, this includes watching at least the first 30 seconds of video. You may need to watch for a bit longer or need to skip to watch a few seconds of other portions in the middle or end of the video in some cases. It may also be helpful to skim over the video transcript [when available].
- To answer this question, you may consider the potential impacts on the lives and well-being of the **public** or **specific audiences** who see the content, and on **the individuals, groups, or organizations described in the content**.

If this claim turns out to be wrong, would it impact someone's

(Select all that apply)

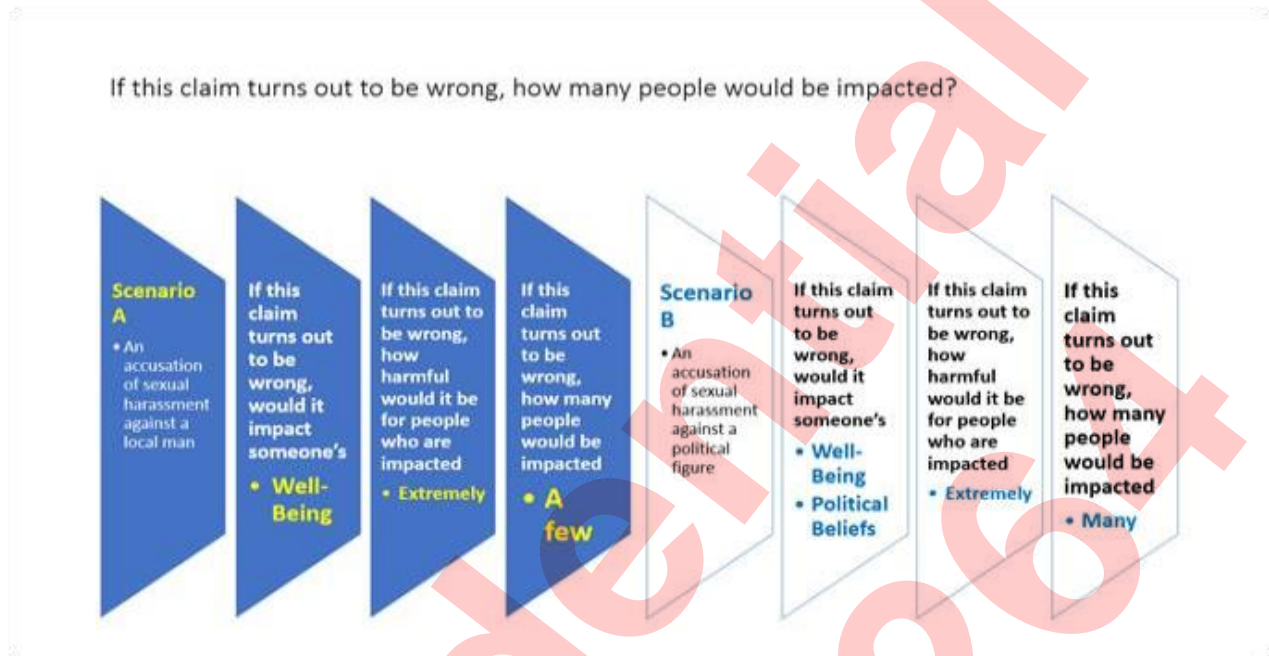
Health, safety, or well-being	Political beliefs or behavior	Views of other groups, such as race, ideology, or religion	Views of the government, the news media, or other institutions	Something Else
<ul style="list-style-type: none"> • Information about Covid • Medical Remedies • Legal Information • Financial Information 	<ul style="list-style-type: none"> • Voting Information • Claims about political figures • Information on candidate stances or party platforms 	<ul style="list-style-type: none"> • Claims about identity groups • Identity groups can be defined by characteristics such as race, ideology, religion, sex, gender identity or expression, citizenship, national origin 	<ul style="list-style-type: none"> • Claims about government • Banks • The healthcare system • Political processes • The education system • The news media • Claims about other institutions 	<ul style="list-style-type: none"> • The claim (in part or full) does not fall into one of the other categories but could impact people who believe it is true (whether it is true or not)

If this claim turns out to be wrong, how many people would be impacted?

- a) Most people
- b) Many
- c) Some
- d) A few
- e) Almost no one

- The thing to consider answering this question is to think about your answers to the impact type and level of harm questions and ask yourself whom it would harm if it were wrong. Keep in mind that something that could be extremely harmful may only impact **a few** people, and something that would have a very low level of harm could impact **most** people.
 - Sports news about a major trade that turned out to be factually incorrect might have a very low impact as far as harm, but it may impact “**Some**” or “**Many**”.
 - A local newspaper article about a man’s arrest might impact his family’s well-being and is potentially extremely harmful, but the impact may be limited to “**A Few**” people, namely family and friends.
 - Further considerations include the potential reach of the claim. Is it a claim about a well-known business that might get nationwide news coverage?
 - Is it a claim that could impact the well-being of the person it is about AND the political beliefs of the public at large, if so, you might consider that to have a wider audience to impact?

Examples:



If this claim turns out to be wrong, how harmful would it be for people who are impacted?

- a) Extremely
- b) Very
- c) Moderate
- d) Not very
- e) Not at all

To answer this question, consider the impact the claim would have, if it turned out to be wrong, on the public, specific audiences who see the content, and on the individuals, groups, or organizations described in the content.

Note: Although we provide some examples of types of claims that often fall into each impact-level category, these are only meant to serve as illustration and are not meant to be strict guidelines for which topics should be assigned to which impact levels

- You should use your best judgment to determine the impact level of each piece of content you evaluate.
 - e.g., while medical claims that might impact a person's healthcare choices would be **extremely** harmful if wrong (e.g., a claim about a new cancer cure), other types of medical claims (e.g., a claim that eating too much chocolate causes acne) might be **moderately** or **not very** harmful.
 - If you are rating a multi-part claim, you should provide the impact rating for the claim or part of the claim that seems to matter the most (i.e., the highest impact or most severe claim).

Examples:

If this claim turns out to be wrong, how harmful would it be for people who are impacted?

Extremely	Very	Moderately	Not Very	Not at All
<ul style="list-style-type: none"> • Medical cures • Voting information • Accusations of criminality • Financial Information • Political information about candidates or parties 	<ul style="list-style-type: none"> • Natural remedies for minor conditions • Accusations of moral wrongdoing 	<ul style="list-style-type: none"> • Local non-crime news stories 	<ul style="list-style-type: none"> • Sports news • Factual claims about items like the weight of an elephant, or the size of a rainforest 	<ul style="list-style-type: none"> • Human interest stories (if not about something more serious) • Entertaining stories about animals

The Satire and Opinion Labels

While you may have to consider what might be purely opinion or obvious satire in the Task B queues as you consider if the content has a central claim, there are specific questions about satire and opinion in Task A.

Satire

Is this content a satire?

a. Yes, it is satire?

i. Which better describes this satire content?

1. Likely to mislead. Some people might mistake the central claim as factual information.
2. Not likely to mislead. Most people would recognize the central claim obviously sarcastic, ridiculously exaggerated, or too absurd to be true.

We ask you to tell us whether the content you are evaluating is Satire:

Satire refers to humorous content that uses irony, sarcasm, or ridicule to expose, denounce, or make fun of human foolishness, immorality, and/or shortcomings.

- The main purpose of satire is humor and entertainment, not to convey facts.

The Likely to mislead and not likely to mislead options for satire posts

If you determine the content is satirical and use the **Yes, it is satire label**, we ask you to consider if the average person might think there is a serious claim embedded in the sarcasm or ridicule. Please note, using the “Likely to mislead” option does not necessarily mean that people are unlikely to recognize the sarcasm, it simply means that there might be a claim that qualifies for evaluation within the content.

Note: Remember, labeling as “**Not likely to mislead**” disqualifies the content from further evaluation in Task B, but the Likely to mislead label passes it through for evaluation.

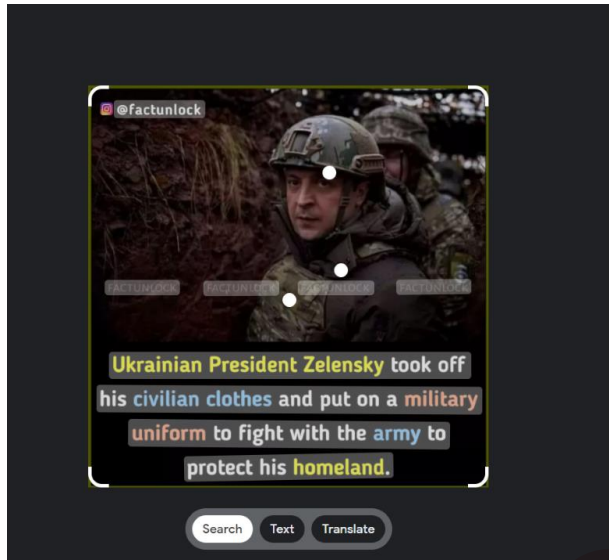
No, not all images, or even memes or those with drawings or overlaid text are labeled as satire, which precludes them from a final evaluation.

Often new raters, or even seasoned ones, misunderstand the satire label resulting in a lot of missed claims and final evaluation mistakes. Satire can be a tricky thing to recognize, but let's start with a few examples that are **NOT** satire and need to be rated. Frequently, the confusion arises on jobs with photos.

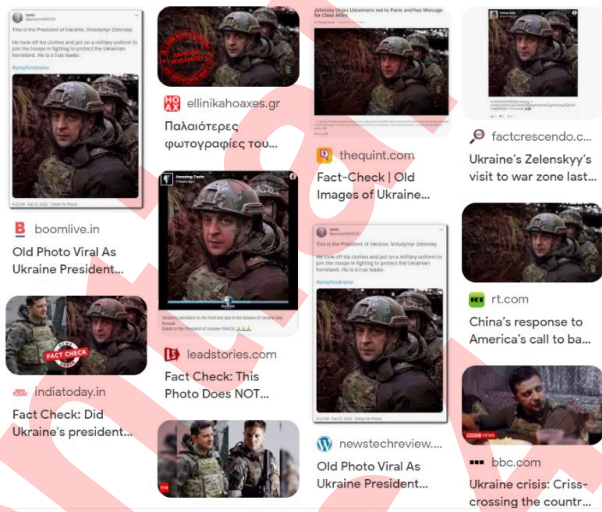
Example:



The example above, despite the overlaid text on the photo, is not satire. The content has a serious claim to be evaluated that Zelensky has joined in the current conflict. A reverse image search brings up evidence that states the claim is not supported.



Visual matches



Example:



Here, even with no caption, you may strongly suspect the image is altered. Even so, an image like this has an implied claim that the photo is authentically Trump, and other former or present world leaders are aggressively surrounding Putin. Even if you think you will not be able to find evidence, the claim should still be identified, and the job should be evaluated. Point being, this should not be labeled as satire, which immediately stops the rating process before the final evaluation. Evidence (found by a reverse image search) shows the photo was indeed altered, making the claim here **not supported/fully contradicted**.

Remember that satire refers to ridiculous or highly exaggerated content that the average person would not mistake for a true effort to pass on factual information.

Hypothetical example:



To help you understand the difference between the above example (labeled as satire) and a photo with a claim that SHOULD BE evaluated (not labeled as satire) please consider the next contrasting hypothetical example:



While the above is clearly humorous and has a similar sentiment to the image rated as satire, there is also an implied claim here that someone purchased a billboard that states birds are not real. The billboard's authenticity is something that we can evaluate along with an intent to inform (albeit in a humorous way) the reader of the existence of such a billboard.

Note: Remember that to get a satire label the answer to the “Does this content have a central claim” question must be – **Yes, this content has a central claim**. Thus, content that is simply sarcastically funny that is not making ANY claim at all, not even an exaggerated or sarcastic one, will have already been labeled as not having a central claim which completes the Task A labeling by disqualifying the content for further evaluation.

Opinion

We also ask whether the content you are evaluating is Opinion:

- Opinion here means someone's opinion as written in any of the components.
- Some content may contain both opinion and factual statements (e.g., content may contain a factual central claim to support an opinion.)
 - Not all personal webpages, blog posts or social media posts contain opinion. You should only label content from these sources as opinion if it is sharing an opinion.

Opinion posts often have the intent to persuade. Those that ONLY set out to persuade us to the poster's way of thinking and that do not have an intent to inform us to sway our opinions are submitted before this point in the labeling, so the types of posts we must choose between the yes, this is opinion and the no, it does not appear to be opinion are those that have relevant factual claims that attempt to INFORM us to persuade us. In other words, there must be factual claims that we can evaluate as supported or not supported related to the main purpose of the post.

Tip: Remember that ONLY type of opinion posts that get the Yes, it is opinion label are those that have some factual claims to support the opinion, otherwise, the job will be submitted when you selected the “No, this content does not have a central claim” label.

Will NOT Reach the Opinion Question in SRT Flow



Does the content contain a central claim?

Yes, the content has a central claim.

No, the content does not contain a central claim.

No - the content does not contain a central claim. The content states an opinion but does not use statements of fact to support that opinion, providing there is no claim in the accompanying status update.

Claim Location Label

If the content does have a central claim, you will see a question asking where the claim is located, meaning the media or the post text or is it both.

Where is the central claim located in this content?

- a. Text from post author only.
 - b. Media only.
 - c. Text from post author and media separately.
 - d. Text from post author and media together.
- **Text only:** The central claim is in the post text and not in the media (photo, video, and/or link). This includes a text post with no attached media OR a text post with attached media, where the media do not contribute to the central claim.
 - **Media only:** The central claim is in the media (photo, video, and/or link) and not in the post text. This includes an image, video, or link post with no accompanying status message or caption (e.g., a meme) OR images and videos with overlaid text when there is no accompanying post text.
 - **Text and Media Separately:** The central claim is **independently** located in both the post text and the media (photo, video, and/or link), so you don't need both components to identify the central claim. If the components were separated, you would still identify the same central claim for each component.
 - **Text and Media Together:** To identify the central claim, it **requires reviewing both** the post text and the media together. If the components were separated, you would no longer be able to identify or evaluate the same central claim.

More on Evidence Labels

Guidance regarding how to search:

- We ask you to find evidence for or against the central claim you've identified from a trustworthy source.
 - By evidence, we mean a statement, an article about, a fact-checker article, photo, video, audio, or statistic relevant to the central claim.
 - This evidence should be reported by a different source than the one whose content you are investigating.
 - This evidence can either support the initial claim or go against it.

Evidence Dos and Don'ts

Double check the following:

- After searching, you should click into specific pages in the search results.
- Select the source with the highest quality evidence to evaluate the claim in the next steps.
***Note: please make sure the source is discussing the same video/photo as the one you're evaluating, not another similar video/image.**
- You should **look for sources that provide *additional* information or reporting about the video/photo and claim** (for example, sources that describe what is shown in the video/photo, when or where the video/photo was taken, whether the video/photo has been altered/digitally manipulated, and/or whether the claim is supported)

***Note: You should not use a link to an identical copy of the video/photo content you are evaluating that includes no additional information/reporting as an evidence source.**
- **You should check to be sure evidence you use is not describing a different place, person, or event than the content you are evaluating.**
 - e.g., for a claim warning about a series of crimes in a community, check to see whether the evidence speaks to the same community and time as the claim you are evaluating).
- You should check to **be sure evidence you use does not contain facts or statistics pertaining to a different time** than the content you are evaluating.
- That it speaks directly to the same event, time, or video/image as the original claim (Unless the evidence you find is showing that the video/image is of a different time, place or person than described by the claim).
- Meaning, check to be sure evidence you use is not outdated or about previous events, especially for claims about current events.

Inappropriate Evidence

The following ARE NOT appropriate evidence sources:

- The same content you are currently evaluating—either from **the same source (if the source is identifiable) or an identical version of the content you are evaluating from another source** (e.g., you may not use a tweet or Facebook post of the same photo and message you are evaluating as an evidence source).
 - Your evidence **must** include additional reporting/information about the content and its central claim.
 - it is okay to use an evidence link that includes the same video/image you are evaluating **from a different source/website, if it has additional reporting**, facts, or information to support or contradict the claim, beyond what was in the content you are evaluating.
- Other content (articles, links, posts, etc.) from the **same source** as the content you are evaluating (if the source is identifiable, e.g., from a logo on the photo or URL in the content text)
 - You may be able to identify the source of the video/image through logos or watermarks on the video/image.
- **The url for your full search engine results**
 - (e.g., <https://www.google.com/search?q=ice+disk+Westbrook+River>)
- **Social media posts, unless:**
 - you are rating a claim about whether a person said something on social media, and you are providing a link to the post referenced in the claim (however, because people may delete social media posts, a claim shouldn't be labeled false only because the tweet/post is no longer present).
 - the post is from the official account for a person or organization who is an authority on the claim's subject (e.g., a post from local emergency responders' office would be an appropriate source for a claim about a local emergency in their district).

Please note that after the Twitter verified account changes, we no longer consider ANY Twitter evidence as acceptable. You may continue to use other verified social media.

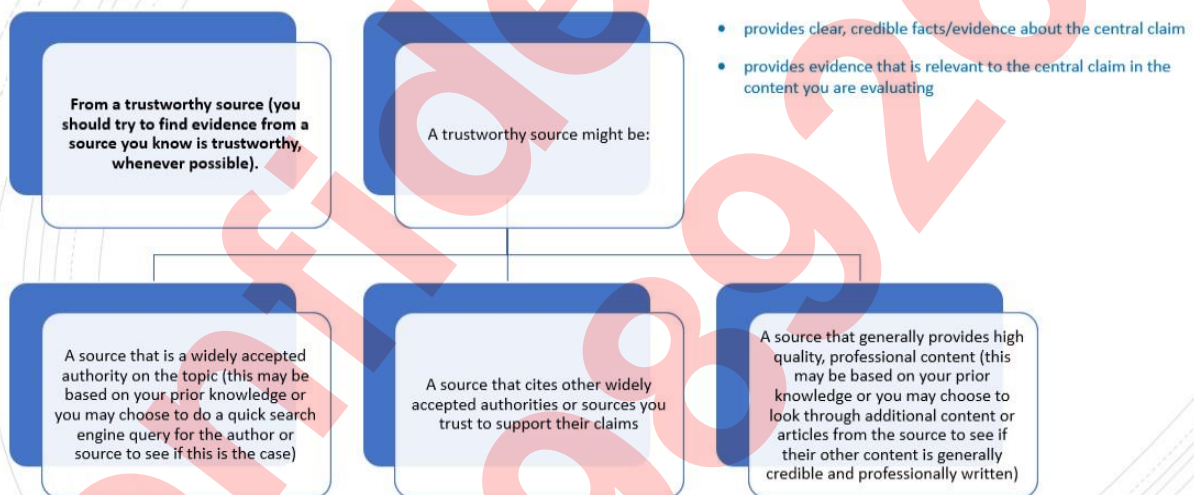
Trustworthiness Labels

Does this evidence come from a source that seems trustworthy?

- a. **Yes**, this source seems trustworthy.
- b. **No**, this source does not seem trustworthy.

Oftentimes, there is some confusion on what makes an evidence source trustworthy. Keep in mind, ultimately, that is up to you, but to help clarify, we would like to give some general guidance and tips.

Here, we ask you to describe the highest quality piece of evidence you found about the central claim in your search. High quality evidence is evidence that is:



**** You are answering this question about the evidence source you find to evaluate the job, not the trustworthiness of the content source of the job itself. ****

If you encounter a job where you think the linked article (or video) comes from a questionable source, this has nothing to do with the trustworthiness of the evidence you find to evaluate that job or to answer the SRT question. For example, if you are labeling a job that links to a YouTube video, and you know that in most cases, social media is not an acceptable source, the trustworthiness of your evidence has nothing to do with the original source of a job.

Please make every effort to keep personal biases out of your evidence assessments. Remember,

your quality is based on labeling jobs uniformly overall in a broad sense with your fellow contributors. Fringe politically biased sites are typically not ideal sources of evidence and most people will consider mainstream news sources with professional reporting and editors to be reasonably trustworthy sources.

More on Final Evaluations

Please recall earlier in the document where we outlined how to align your final evaluation with the claim and the evidence you found.

When you label a claim fully supported, SRT moves you to the final questions and the submit button; however, if you select partly supported or not supported, there are more questions to answer about your final evaluation.

Please see the table below for characteristics of the follow up labels:

Fully Supported	Partially Supported	Not Supported
The central claim is fully supported by evidence from one or more trustworthy sources. I found evidence from one or more trustworthy sources that supports the main aspects of the claim, and no major aspects are contradicted.	Evidence from a trustworthy source only partially supports the central claim. This may be because: <ul style="list-style-type: none"> Major parts of the claim are supported, but other major parts are misleading or out of context. Major parts of the claim are supported, but there is no evidence for other major aspects of it. The claim is about something someone said, and the quote is correct but is taken out of context or has been significantly misrepresented. The claim is represented in significantly different ways in the headline or title of the article/link vs. body of the article/link itself you are evaluating, and not all aspects of the claim are supported by evidence. 	I found evidence from a trustworthy source that contradicts (but does not support) some or all of the claim, OR I found one or more trustworthy sources that concludes there is no evidence to support the claim.
If the claim is fully supported, move to the next question. Move to Question 12	If the claim is partially supported, further questions open for the evaluation section	If the claim is not supported, further questions open for the evaluation section
	Based on the evidence you found, is the central claim: <ul style="list-style-type: none"> Partially contradicted Part of the claim is supported, but other parts are contradicted, misleading or out of context. Not contradicted Parts of the claim are supported, but there is no evidence for other aspects of it. 	Based on the evidence you found, is the central claim: Fully contradicted I found evidence from one or more trustworthy sources that contradict the main aspects of the claim. Partially contradicted Some parts of the claim are contradicted, some parts of the claim not addressed. Not contradicted The trustworthy source addresses the central claim, but ultimately concludes that there is no evidence to support or contradict the central claim.

Furthermore, if your final evaluation is Partially or Not Supported, after you answer the questions about the degree of contradiction (fully, partially, or not), you will also be asked for the scenario that is preventing the claim from being fully supported.

Based on the evidence you found, do any of the following describe this content? (Multiselect)

- a. Misleading headline
- b. Out of context
- c. Digitally altered media
- d. Scam

The content has a misleading headline/title	The headline/title of the content you are evaluating states an exaggerated version of the claim that is not supported by evidence, but a less extreme version of the claim stated in the body of the content is supported by the evidence that you found.
Out of Context	The accompanying media is of a different person/place/object/time/event than what is described in the central claim.
Digitally Altered Media	the photo or video has been digitally manipulated/altered/morphed to show something different than the original photo or video
Scam	The content is a link to coupon, giveaway, offer, or promotion that the evidence you found indicates is fake or fraudulent (especially one that requests that people provide information to receive the offer).

Tips and How to Avoid Pitfalls

Writing claims like a pro

What should you **not do** when forming or writing a central claim?

- Don't ever form your central claim such that it is only one or two words or is just a person's name (that is an incomplete claim).
- Don't write "opinion" for the central claim if you think the job contains opinion.
- Do not use a claim such as "news" or "covid".
- Do not use your evidence link as the claim.
- Do not use your final evaluation as the claim, such as "it is false or not supported that the sky is green".
- Do not write "agree" or "disagree" for the central claim (that is not a claim).
- Do not write "The claim is". Only type in what the central claim is. There is no need to begin the sentence stating what the claim is.
- Do not copy and paste the entirety of the post text (edit and simplify, remove emojis, any unnecessary information, make the claim as uncomplicated as possible).
- Do not paste any links.
- Do not copy the transcription text of a video.
- Do not copy the entirety of the text contained within linked media.
- Do not ever let your claim be affected by how easy or difficult it is to research.
- Do not leave the central claim section blank if you have identified a central claim from the post for evaluation.

Dos for writing a central claim

- Use your own words.
- Be specific.
- Make sure that your claim encompasses the complete main purpose of the post. For example, when part of the claim includes photo authenticity, be sure to write it – "Russia has invaded Ukraine and this photo shows damage from airstrikes".
- Try to use a complete sentence or a complete main idea.
- Make your claim easy to understand.
- The written claim may be in your market language or in English.

Photo or video is the claim

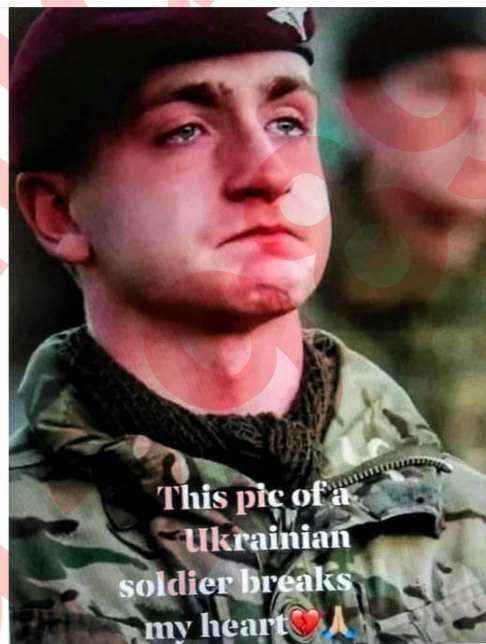
Photos are often part or even ALL the claim

You will notice on Project Uolo that often the content includes a photo. Most of the time, the picture will also include a caption on it or is accompanied by post text that may give some clues on what it means or what it is reported to be. Sometimes, though, the post text or caption may be something as non-descriptive as 'ha-ha' or 'wow'. In those cases, it does not always mean that the job has no claim.

In many jobs in all markets, the photo itself and its authenticity, or that it is what it is reported to be, is all or part of the claim.

Sometimes the claims about the photos will be derived from what the post text or caption says or implies about them.

Content in Review



In the above example, the photo is an integral part of the claim. The overlaid caption is referencing this as something specific. Remember that when you are identifying claims from the content you see in production, every component of the content must be considered, the post text or caption, the media, a linked article, etc. Sometimes, as in the case above, the components will obviously go together (post text and photos in these examples).

Importantly, had you simply labeled as no claim because the claim is mostly contained in the

photos, you would have missed misinformation. The primary objective of the project is to fully evaluate claims on social media as supported, not supported, or somewhere in between, and missing claims will prevent you from meeting this objective.

The trickier photo scenarios

Sometimes, the implication that the photos are related to the post text is much more subtle.

Suppose this image was accompanied by post text that stated, *"Considered one of the most dangerous locations in the world, the San Andreas fault is in western USA's state of California. This crack that is 1200 km in length runs along the north American and Pacific plates."* and evidence does indeed show the statements about the fault line to be true because the post implies that this is an authentic photo taken from there, whether the photo truly depicts that needs to be part of your claim.



Evidence shows the photo is not the San Andreas Fault, but instead is a Faultline in Utah. Again, had you not included the photo as part of the claim, you would have missed those parts of the complete claim (the inclusion of the photo) is contradicted. Partly supported/Partly contradicted is appropriate here.

When might images be the entire claim?

Suppose you encountered the following content without comment or caption:



Even without a caption or post text, the photo itself implies that a massive deforestation has taken place in a 10-year span in this location. In this case, if you do indeed infer that claim, you could try a reverse image search, which does produce evidence that shows this is a faked photograph, as it is the same photo simply split in half to show the progress of clearing an area. Had you labeled as no claim, you would be missing misinformation. The implied claim about the deforestation would be contradicted.