

Annotation Guidelines

Your job is to annotate the relation between the social media *post* and the fact-check *claim*. **Is the claim implied by the post?** You should **emulate a fact checker** who is making the decision of assigning a fact check to a user post. **Try to interpret the post in the way the user intended to.** The annotation schema has three options (Yes - No - ?), and “?” *with meaning that you can not tell.*

- *Yes (Relevant)*
 - **The claim is fully present in the post** or
 - The post seems to be related to the claim, but **some details are different**. The claim might only be suggested.
- *No (Irrelevant)*
 - **The post does not contain the claim at all.** At most they talk about similar topics.
- *? (Can not tell)*
 - *You can not tell because you do not understand the language and the translation is not good enough to make a decision*
 - *You do not understand the topic and you are not sure whether the claim is implied or not. - For this reason, **you can check the fact-check article linked to each fact-check.***

To evaluate the relevance between a social media post and a fact-check, it does not matter what the *stance* of the post is. Even if the post disagrees with the claim, it might still be a relevant pair if they are about the same idea.

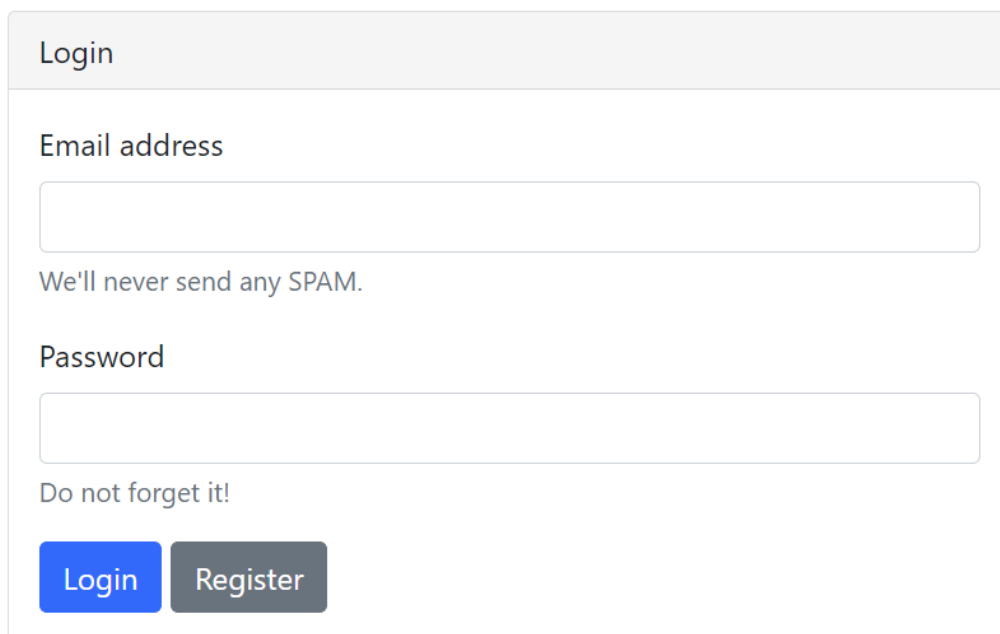
Annotators

The expected time for evaluating one pair of a social media post and a fact-check is approximately 20 seconds, while each document contains multiple pairs of the same post and different fact-checks. Our task is to evaluate the relevance between each fact-check claim provided and the social media post, which is on the top of the web page.

Registration and annotating data

First of all, it is necessary to register into the annotation tool, which can be accessed on this URL.

1. **Registration** - To register, you need to input an email address and password. Then click on *“Register”*. It is necessary to remember the email and password because these credentials will be used to log into the system using the same user interface.



The form is titled "Login" in a light gray header. Below the header, there are two main sections. The first section is for email address, with the label "Email address" and a text input field. Below the input field is the text "We'll never send any SPAM." The second section is for password, with the label "Password" and a text input field. Below the input field is the text "Do not forget it!". At the bottom of the form, there are two buttons: a blue "Login" button and a gray "Register" button.

2. **Login** - To log into the system, provide your email address and password and click on *“Login”* button.
3. After logging into the system, you will see the interface below, where you should click on *“List of documents 2”* to get into the list of all social media posts that should be

annotated

User

ivan.vykopal@kinit.sk

Logout

4. The list of all social media posts for annotating is shown below, where
- Document ID** is the ID of the file from our annotation tool, and in [#Annotators](#), we refer to this Document ID in the column “Range of document IDs”.
 - File** contains the name of the file (mostly the id of social media post from our dataset)
 - Status** - the information how many fact-checks were already annotated
 - Detail** - link for annotating the selected social media post and the pairs between social media post and retrieved fact-checks

Document ID	File	Status	Detail
1	21217		→
2	21306		→
3	8661		→
4	8538		→
5	6687		→
6	7521		→
7	21320		→
8	8463		→
9	9076		→
10	9130		→
11	8378		→
12	9614		→
13	552		→

5. Firstly, all annotators are supposed to annotate the first 12 Document IDs as noted in [#Annotators](#), which can help us to identify correlations and alignment between human annotators
6. Then you are supposed to annotate the data based on the information provided in the table in [Annotators](#) and specifically to annotate the Document IDs from your selected range, e.g, if the range is from 312 - 411, you are supposed to annotate the social media posts with Document ID starting with 312, until the Document ID 411 including. After you finish the annotation, please mark the information in the table

above and if you still have free capacity, please select another range of IDs that is not already occupied.

Example of the Annotation tool

The screenshot displays the Annotation tool interface. It is divided into two main sections: 'Post 8023' and 'Fact check claims'.

Post 8023: This section shows the original text from a social media post in Czech and its English translation. Arrows point from the text to labels: 'Exact text from the social media post' points to the Czech text, 'OCR from images within the social media post - Text extracted from images' points to the English text, and 'Translated texts into English' points to the English text.

Fact check claims: This section lists fact-check claims and their titles. Arrows point from labels to specific entries: 'Claim that was verified by the fact-check article' points to the first claim, and 'Title of the fact-check article' points to the first title.

Claim	Title	YES	NO	?
Claim: Ursula von der Leyen označila seniory za nepřátele EU	Title: Ne, předsedkyně Evropské komise neoznačila seniory za nepřátele EU			
Claim: Klaus Schwab napsal, že pandemie vyhladí 4 miliardy lidí do roku 2050	Title: Text podporující „vyhlazení“ 4 miliard lidí nepochází z knihy zakladatele Světového ekonomického fóra			

A fact-check **claim** and fact-check **title** can contradict each other since the title is written by fact-checkers, and in most cases this **title** already tries to debunk the false information spread. Sometimes the **title** can differ from the debunked **claim**, as some fact-check articles address multiple false claims, and then the **title** is more general or sometimes unrelated to the claim.

Fact-check articles for each social media post were obtained automatically from our database using Information retrieval (IR) tools (especially using text embeddings and calculating cosine similarity between social media posts and fact-check claims). Therefore, it is very often that fact-check claims are unrelated and irrelevant to social media posts and may be from a completely different topic since the IR system is not perfect.

Process

- **Reading a social media post** - In the annotation tool, you will see the text from the social media post. Please read it carefully to understand the main claims spread by the social media post. Your annotation should consider only textual information, so in the case, if there are some images, consider only text for your annotation.
- **Read a fact-check claim and title** - For each social media post, there will be multiple fact-checks, but the number of fact-checks differs for each social media post.

Please, read carefully the claim and title and based on the social media post and the claim and title, annotate the relevance between fact-check and social media post as mentioned above. By selecting one of the options (Yes, No, ?) you will annotate the specific pair of fact-check and post. After providing annotations for each fact-check, click the **SUBMIT** button at the bottom of the page.

Examples

Example #1

In the post below the text does not say anything about Twitter, but this is not the core of the claim. Note that in this case the full post is clearly full match because the image is a Twitter screenshot.

Post:

- **Text:** The Ukrainian hotel Mir, which, according to CNN journalists, was mercilessly bombed by the Russian armed forces. Oh no, wait. These are the Serbian Zvonačka-Banja spas, which are awaiting reconstruction. By the way, they are looking for an investor. Please send CNN a thank you note for the ad. Link to the video of the Serbian spa and hotel: <https://www.youtube.com/watch?v=oe-0pAuSbNE>
- **OCR:** CNN CNN Breaking News @cnnbrk Majk Majkonski These are the remains of the once famous ukrainian resort hotel Peace after the air attack by russian bombers #Standing With Ukraine

Fact-Check:

- **Claim:** CNN posted a fake news story on Twitter about the bombing of the Mir hotel in Ukraine
- **Title:** The alleged tweet, according to which CNN was supposed to report on the bombing of a hotel in Ukraine, was falsified

Expected Label: Yes (Relevant)

Example #2

In the post below, the text is about how Ukrainians behave in trains, but the post does not say that they are making a mess. That is a significant detail of the claim that is not present.

Post:

- **Text:** Thanks Ukrainians! Slovaks, wake up already! However, the reality is different and whoever wants to know it spends that little bit of time to get to know how it really is and what the real patriotic Ukrovs actually want and what they are capable of. As long as you have a strong stomach for it. It is a nation that supposedly belongs to Europe. Czech railway workers probably have a different opinion when they welcome migrants in the express train from Košice to Prague without payment and without

masks, with backpacks full of food donated by our citizens. Keep sending because our TVs can't do it.

Fact-Check:

- **Claim:** Photos show the mess and dirt left on the train by Ukrainian refugees
- **Title:** Old photos from 2015 were used to defame refugees from Ukraine

Expected Label: Yes (Relevant)

Example #3

In the post below I chose irrelevant because it was not clear what the user is talking about without seeing the image. Buča was not mentioned and photo manipulation is only suggested. However, with a full post it becomes clear.

Post:

- **Text:** Since it's probably in Ukraine and Edko has the way of thinking he does, this photo was published on page 148... So that I don't get locked up, I won't write the date 🤔🤔😂😂😂😂 Everyone takes pictures in the same place as on the carousels 🤔🤔😂😂😂 Maybe they only mean the head.... 🙄🤔 It's strange that he has a helmet on the train and only that stupid head on the battlefield 🙄🙄😂😂😂

Fact-Check:

- **Claim:** The pictures prove that Heger faked his visit to Buč
- **Title:** The fabricated photo montage was used to question Heger's visit to Buč

Expected Label: No (Irrelevant)

Example #4

In the post below, both 2008 and AstraZeneca are mentioned in the post and OCR. However, it is not implied by the text alone that it is AstraZeneca vaccine that was made in 2008. This becomes clear when we take the image into consideration, because now it seems that the user is talking about the vial with vaccine with 2008 written on it. We need more than mentions of certain data to choose the relevant label.

Post:

- **Text:** According to the Slovak, this is also a mistake. Made in 2008. Still nothing
- **OCR:** 14:02 15:17 @QWWG1WGA 16PLUS1 AstraZeneca AB0003 05-2021 Arpi Elias VČERA O 23:27 3041572008 ||| ** | 80% q : COVID-19 Vacc injection COVID-19 vaccine (ChAdOx1-5 intramuscular Multidose via AstraZeneca vaccine Vaccine delivery to federal states et 10,485 new infections and 689 new deaths Corona pandemic: claims for Zen

Fact-Check:

- **Claim:** The photo proves that the AstraZeneca vaccine was produced in 2008

- **Title:** Image shared by thousands isn't proof AstraZeneca vaccine was made in 2008

Expected Label: No (Irrelevant)

Edge cases

Example #1 - Noise in OCR

The OCR we use is quite noisy. You can ignore this noise, as long as it doesn't interrupt anything important. If the noise breaks a sentence, do not take the meaning of the sentence into consideration when you evaluate *the relevance*.

OCR noise is okay in the example below. *"These are the remains of the once famous ukrainian resort hotel Peace after the air attack by russian bombers"* is a valid text that you might take into consideration. There is some noise before and after (hashtag, user handle, etc.) that you can ignore.

Post:

- **Text:** The Ukrainian hotel Mir, which, according to CNN journalists, was mercilessly bombed by the Russian armed forces. Oh no, wait. These are the Serbian Zvonačka-Banja spas, which are awaiting reconstruction. By the way, they are looking for an investor. Please send CNN a thank you note for the ad. Link to the video of the Serbian spa and hotel: <https://www.youtube.com/watch?v=oe-0pAuSbNE>
- **OCR:** CNN CNN Breaking News @cnnbrk Majk Majkowski These are the remains of the once famous ukrainian resort hotel Peace after the air attack by russian bombers #Standing With Ukraine

Fact-Check:

- **Claim:** CNN posted a fake news story on Twitter about the bombing of the Mir hotel in Ukraine
- **Title:** The alleged tweet, according to which CNN was supposed to report on the bombing of a hotel in Ukraine, was falsified

Expected Label: Yes (Relevant)

Example #2 - Noise in OCR

OCR noise is not okay in the post below. *"98 14:34 14:39 INT"* interrupted the sentence shown by the doctors and this sentence should be disregarded.

Post:

- **Text:** Please do not get vaccinated
- **OCR:** Please u leaves 98 14:34 14:39 INT you Not vaccinate

Fact-Check:

- **Claim:** German health workers used a photo to discourage people from vaccination

- **Title:** German health workers did not warn against vaccination, this photo was manipulated

Expected Label: No (Irrelevant)

Example #3 - Claims about photographs / videos

Claims are often made about photographs and videos. It is still a good match even when the post itself does not mention the photograph or video and only a description is provided.

E.g. in the post below the text does not mention any video, but the description matches the description of the video from the claim. This should be evaluated as a relevant pair.

Post:

- **Text:** Copied from snift... ⚡ Finland is moving military equipment to the border with Russia. There are several dozen Leopard tanks, armored infantry vehicles and other armored vehicles on the railway sets. 😡

Fact-Check:

- **Claim:** The video shows Finland moving tanks to the eastern border with Russia
- **Title:** These tanks are headed for a military exercise in western Finland, not the Finnish border with Russia

Expected Label: Yes (Relevant)

Example #4 - Grammatically incorrect posts

Sometimes the users write posts that are not grammatically or semantically correct. In that case, try to interpret the post in the way the user probably meant it

E.g. in the post below, the user probably wanted to say Media "presents a tornado as Ukraine" but the user probably did not use the word "presents" correctly. In this case I would also consider using the relevant label because the post does not mention ČT or their intentions .

Post:

- **Text:** THE MEDIA PRESENTS UKRAINE as a TORNADO in MIKULČICY!!!

Fact-Check:

- **Claim:** CT deliberately used a photo of the damage after the tornado to illustrate the war in Ukraine
- **Title:** Czech TV mistakenly used a photo of the damage after the tornado in Moravia in a report on the war in Ukraine

Expected Label: Yes (Relevant)

Example #5 - Using background knowledge

You can use common background knowledge to evaluate the posts.

E.g. in the post below, the post does not say that "Russians" captured the general, but we can assume that this detail is implied.

Post:

- **Text:** NATO commander General Cloutier was captured near Mariupol. NATO Lt. Gen. Roger L. Cloutier, who trained the Azov Nazis, ended up with them in Azovstal and was captured.

Fact-Check:

- **Claim:** The Russians captured NATO General Cloutier near Mariupol
- **Title:** No, the Russians did not capture a NATO general in Mariupol

Expected Label: Yes (Relevant)

Example #6 - Using background knowledge

On the other hand, the post below is trickier. There are hoaxes about all kinds of vaccinations and without dates, it is impossible to tell what vaccination is being talked about.

Post:

- **Text:** The vaccinated cannot even donate blood. So it's basically frozen already!!

Fact-Check:

- **Claim:** A person cannot donate blood after being vaccinated against Covid
- **Title:** You can donate blood after the Covid-19 vaccine after 2 to 4 weeks, similarly to other vaccinations

Expected Label: Yes (Relevant)

Example #7 - Time sensitivity

Evaluate the posts based on whether the match was correct at the time they were written

E.g. in the post below "von der Leyen" is referenced only as the president of the European Commission. This is true now, but it is a time-sensitive fact.

Post:

- **Text:** i Friends, how dare the President of the European Commission say that the new enemy of the EU is the elderly!!! I've had enough of delusions and sick conclusions of the whole EU! The road does not lead here and it only confirms that the EU is a path of self-destruction!!! 🙅🏻🙄

- **OCR:** ONLINE: Corona virus in the Czech Republic and in the world Brussels has a new enemy - the elderly. He wants to completely push them out of everyday life. European Commission President Ursula von der Leyen said that the new enemy of the EU is the elderly. These old people are said to be nostalgic for the good old days", undermine the integrity between EU member states. Seniors claim that during the existence of individual nation-states it was better. In this way, they mainly negatively affect the youngest generation of their descendants, to whom they communicate this false information. This has a serious impact on the upcoming the emergence of a new European state. In this context and with regard to the current pandemic, Ursula von der Leyen insists that all seniors in Europe must remain in isolation at least until 2021. This is the only way to break through all family ties with all the negative consequences for the EU. Brussels 14/04/2020 10:15 Approx Com EarmeeCare European Commission European Commission Common How European

Fact-Check:

- **Claim:** Ursula von der Leyen described senior citizens as enemies of the European Union
- **Title:** Seniors are not the enemy of Brussels. It is a hoax

Expected Label: Yes (Relevant)

Example #8 - Conjunct claim

Sometimes the claim might in fact contain multiple claims. In that case, it is enough if the post implies at least one of them.

E.g. in the post below, there are actually two claims being made: (1) Nicholsonova was living as a prostitute, and (2) Nicholsonova protested for sex workers' rights. However the post only implies (1) "Nicholson was supposed to make a living as a prostitute in Canada" and (2) is not mentioned at all. In this case, it is still a relevant pair.

Post:

- **Text:** Lucia Ďuriš Nicholsonová should have been arrested in Canada for prostitution in 2003.
- **OCR:** OTTAWA POLICE SERVICE SERVICE DE POLICE D'OTTAWA king for Executive director, Ottawa Police Service Board TO/DEST FROM/DXP Chief of Police, Ottawa Police Service SUBJECT/OBJECT: European citizen/Slovak female C 13-4654/13-9876 DATE/TIME 12-18-2003 1900-2300 LOCATION: Belleville and Commwall run The end DISTRIBUTION Ongport Com CA COP, Caps P City Cm, City M TYPE OF INCIDENT OPERATION CROSS COUNTY SEX TRAFFICKING OPERATION SUMMARY OF CRIME INCIDENT and the th th of the O During one of the undercover investigations, offic NICHOLSON, Second Surname KUBOVICOVA, cit FOUND) REPORT RAPPORT On 12-18-2001, the South Ottawa Police Department participated in the Operation Cross Country which is part of the Bureau's innocence isst National ative to combat juveline sex trafficking The goal of the national operation is to identify and rescue victims of sexual and human trafficking and to arrest the

predators responsible for victimizing these children and young adults. During the undercover operation South Ottawa police officers contacted and arrested escorts who were soliciting for prostitution on different prostitution based we officers booked appointments them The female (NICHOLSON), according to witnesses, showed signs of bel apartment complex near The Brass Club (Private Erotic Massages for Adults agent posing as a client booked an appointment and was directed to the location. On apartment, officers contacted two female escorts in the sparse two-bedroom unit. Evidence o brothel type operation, was found in the apartment as well currency from the operation which seized as evidence. The escorts pimp was later contacted and arrested at the scene. The investigation is continuing to ascertain whether the females were being sexually trafficked as part of a larger prostitution ring spanning across several Ottawa cities and counties Co poslankyne? Nicholsonová sa mala živiť ako prostitútka v Kanade. All of the escorts contacted or arrested were provided support, alternatives, resources and direction to victim service agencies in Ottawa County. Kur.a je to, ja ji znám, ja s ní žiju...

Fact-Check:

- **Claim:** Lucia Nicholson lived in prostitution and protested for the rights of people in the sex industry
- **Title:** Lucie Ďuriš Nicholsonová did not participate in the protest for the rights of people in the sex industry

Expected Label: Yes (Relevant)