



Annotation Guidelines for the ClimateCheck Project

1. Introduction

The ClimateCheck annotation project aims to develop datasets for developing AI models that link social media claims from laypeople to scientific abstracts in the climate change domain. The idea of the project is to facilitate fact-checking processes of claims made online by providing scientifically valid proof from peer-reviewed articles.

As annotators, your job will be to develop a gold dataset (i.e. data annotated by humans) which will be used in training and testing AI models. The data you will create will be used in a shared task (a competition where different teams work on developing models using the same dataset)¹.

As advanced students in the field, your expertise in efficiently reading and comprehending the scientific jargon in research papers is essential in this process. Briefly explained, your task will be to:

- Read a claim made about the climate on social media.
- Read an abstract connected to the claim.
- Determine whether the presented abstract supports the presented claim, refutes it, or does not have enough information (NEI) for you to decide.

Although you might know more about a topic than the presented abstract, the label you choose (supprts, refutes, NEI) must be decided **solely based on the abstract, and not on the additional knowledge you might have about the topic**. This is because we want the AI model to learn how to connect two texts without presenting additional information.

Each annotator will be grouped with another fellow annotator who will receive the same claim-abstract pairs. This will allow us to see the **annotation agreement in the data**, proving that the labels were chosen carefully and are not based on subjectivity. **We aim to obtain a high inter-annotator agreement score**, which is why you must take the time to read the guidelines carefully and make sure you apply them when working.

¹ The shared task will most likely take place with the Scientific Document Processing workshop this August: <https://sdproc.org/2025/sharedtasks.html>. The workshop is co-located with this year's Association for Computational Linguistics (ACL) conference: <https://2025.aclweb.org/>.

2. Using INCEpTION

Downloading and installing

In this project, we will use the annotation tool INCEpTION. For that you will need to:

- Go to the INCEpTION website: <https://inception-project.github.io/>.
- Download the latest INCEpTION version by clicking on the 'Download' box on the right side of the page.
- A step-by-step guide for installation can be accessed here: [Installing and Starting INCEpTION](#).

Accessing our project instance

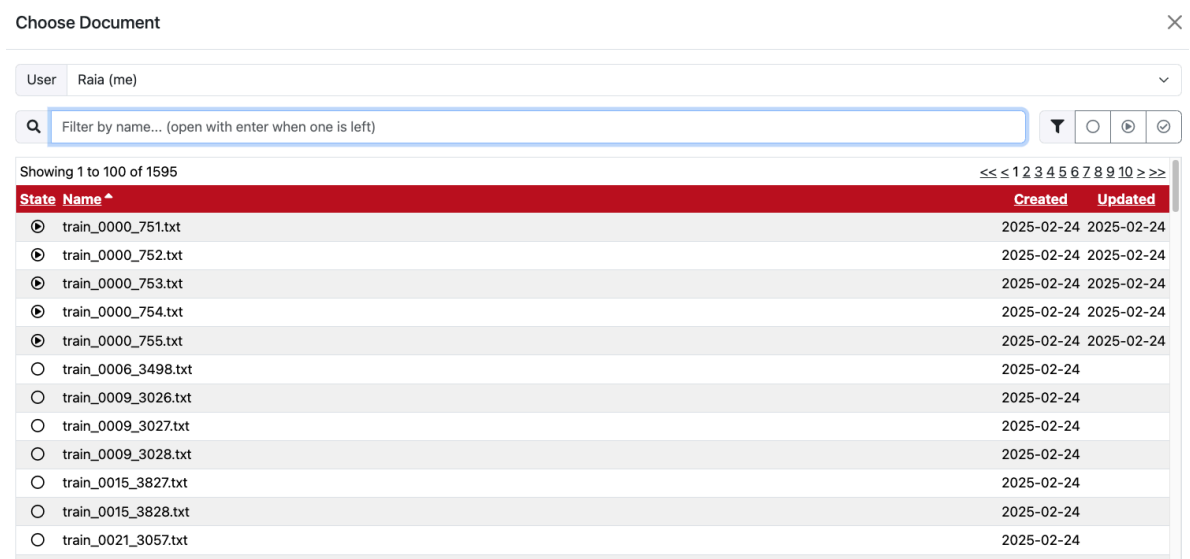
To access our INCEpTION instance, you will need to connect to the DFKI VPN first.

After you are connected, you can access the instance here: <http://172.16.150.205:8080/?2>.

Please use the username and password provided privately to you.



3. Accessing documents to annotate


Once in the Inception instance, choose the relevant project by clicking on it. You will see a left-hand bar, in which you need to click on 'Annotation'. You will be presented with a screen similar to this:



Choose Document			
User Raia (me)			
Filter by name... (open with enter when one is left)			
Showing 1 to 100 of 1595			
State	Name	Created	Updated
▶	train_0000_751.txt	2025-02-24	2025-02-24
▶	train_0000_752.txt	2025-02-24	2025-02-24
▶	train_0000_753.txt	2025-02-24	2025-02-24
▶	train_0000_754.txt	2025-02-24	2025-02-24
▶	train_0000_755.txt	2025-02-24	2025-02-24
○	train_0006_3498.txt	2025-02-24	
○	train_0009_3026.txt	2025-02-24	
○	train_0009_3027.txt	2025-02-24	
○	train_0009_3028.txt	2025-02-24	
○	train_0015_3827.txt	2025-02-24	
○	train_0015_3828.txt	2025-02-24	
○	train_0021_3057.txt	2025-02-24	

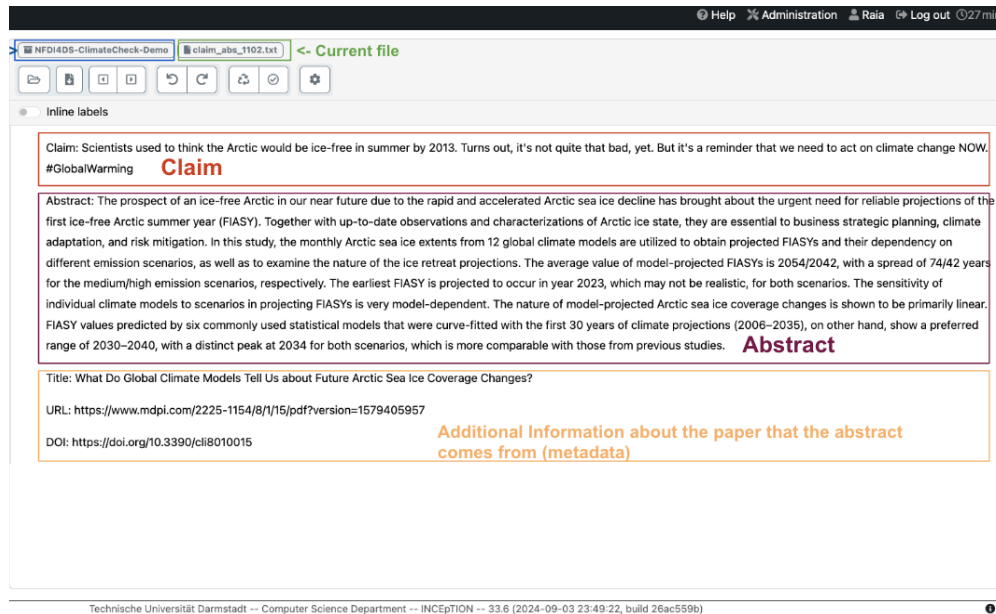
These are the documents that are assigned to you to annotate. Each document can have one of three possible states:


-  : the document is still unopened and unannotated by you.
-  : the document was opened by you and annotation is in progress.

-  : the document was annotated and closed. These documents cannot be opened again unless you contact an admin.

4. How to annotate

By clicking on a document, you will be directed to the main annotation screen, which will look like this:



Additionally, you will be given access to a spreadsheet, which will include all the documents assigned to you. For example, this is a link to a spreadsheet containing the documents of the demo project:  ClimateCheck Demo: Metadata .

Sometimes, the abstract might be difficult to read using the annotation tool, which is why one of the given metadata fields is the URL of the original paper. The easiest way to access the full paper is to copy the document ID, find it in the spreadsheet, and click on the URL there.²

² You can also copy and paste the URL that is presented in INCEpTION but that might be difficult due to how the tool is programmed.

After you read the claim and the abstract, follow the points below:

- Read the claim and ask yourself the following questions:

1. Is it scientifically check-worthy?

A check-worthy claim in the scientific sense means that the information in it can be verified against a scientific article. This, for example, would exclude political claims such as *"Tony Abbott says Australia's per capita emissions targets for 2030 are 'the best in the developed world'"*. **If the claim is not scientifically check-worthy, please note it when you close the document, and don't annotate it.**


2. Is it atomic?

Atomic claims convey a single, clear, indivisible, and context-independent piece of information. **If the claim in front of you is not atomic, try to split it into several atomic claims, and annotate it only based on one part of it. Then, add a comment when you close the document on which part you annotated.**

- Read the abstract and decide its relation to the claim:
 1. **Supports:** if the abstract contains information that supports the claim.
 2. **Refutes:** if the abstract contains information that refutes the claim.
 3. **Not enough information (NEI):** if the abstract does not have enough information for you to decide.

There might be cases where it is difficult to determine which label to give a document. As those cases cannot always be anticipated, we will gather them throughout the project and decide how to annotate them on a case-by-case basis. If you encounter such cases, **add a comment about the document in Inception explaining the difficulty you are facing**. In case the issue is a general problem that might be encountered by other annotators, it will be discussed with everyone and added to this document.

Once you decide on the label, double-click the word "Claim" until it is highlighted like this:

 **Claim:** "natural gas" is considered cleaner than coal and oil

Make sure only this word is highlighted, because if the two annotators highlight different words, it will affect the agreement score.

Once you highlight, a label Selection box will appear on the right. Choose your label and then on the "check" icon to close the document.

The screenshot displays the NFDI4DS ClimateCheck interface. At the top, there's a header bar with the text "NFDI4DS - ClimateCheck - Group 2" and a file name "train_0000_751.txt". Below this is a toolbar with icons for file operations (upload, download, left arrow, right arrow, undo, redo, refresh, checkmark, and settings). A red arrow points to the settings icon. Below the toolbar, there's a toggle switch for "Inline labels". The main content area shows a document with a claim: "Claim: 'natural gas' is considered cleaner than coal and oil". Below the claim is an abstract text: "Abstract: A well-known theorem by Herfindahl states that the low-cost nonrenewable resource must be exploited first. Consider resources that are differentiated only by their pollution content. For instance, both coal and natural gas are used to generate electricity, yet coal is more polluting. We show that the ordering of extraction need not be driven by whether a resource is clean or dirty. Coal may be used first, followed by natural gas, and again by coal. Such 'vacillation' does not occur under cost heterogeneity. A perverse policy implication is that regulating pollution may accelerate use of the polluting resource. (JEL Q32, Q38, Q53, Q58)". Below the abstract is the title: "Title: Ordering the Extraction of Polluting Nonrenewable Resources". Below the title is the URL: "URL: http://publications.ut-capitole.fr/2734/1/ordering.pdf". Below the URL is the DOI: "DOI: https://doi.org/10.1257/aer.98.3.1128". On the right side, there's a sidebar with a "Span" section containing "Delete" and "Clear" buttons. Below this is a "Layer" section with a "Label" input field. Below the "Layer" section is a "Text" section with a "Claim" input field. Below the "Text" section is a message: "No links or relations connect to this annotation." Below this is a "Label" section with a dropdown menu showing "Not Enough Information", "Refutes" (highlighted in blue), and "Supports".

Remember: closing the document means you cannot change the chosen label. Documents can be re-opened only if you request it from an admin.

5. Additional support

If you have any questions or concerns regarding any part of this process, please don't hesitate to contact me either on Microsoft Teams or via email!

Thank you for participating!

Raia