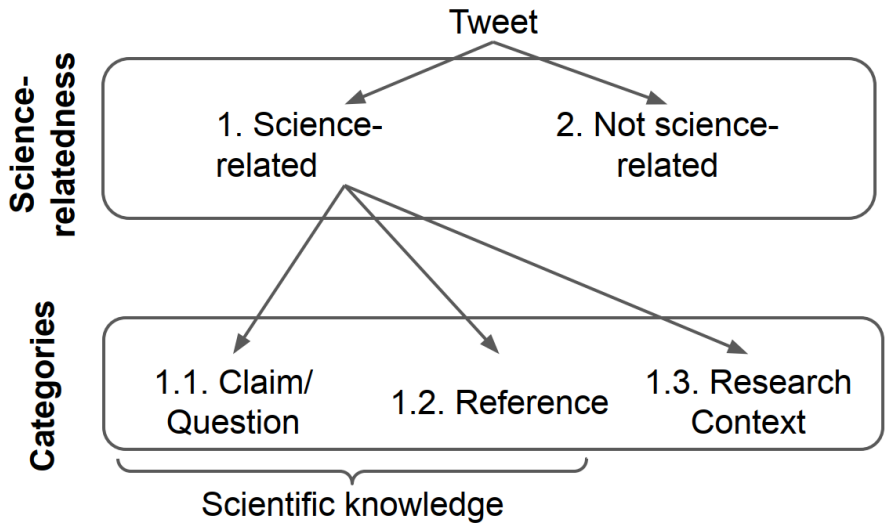
Labeling Framework for SciTweets

Description

This labeling framework includes definitions and labeling instructions for different forms of scientific relatedness of online discourse in Tweets. Science-relatedness is divided into two categories: one positive for science-related texts (**Category 1**) and one negative category for unrelated texts (**Category 4**). For **Category 1** we distinguish three different forms of science-relatedness (see Figure below):

- **Category 1.1**: texts that contain or are meant to contain scientific knowledge
- **Category 1.2**: texts that contain a reference to scientific knowledge
- **Category 1.3**: texts that are related to scientific research in general

All texts that do not belong to **Category 1.1** or **Category 1.2** or **Category 1.3** belong to **Category 2**.



Task

For each text, answer the questions in ascending order starting with **Category 1.1**. If **Category 1.2** is answered with **Yes**, then **Category 1.3** does NOT need to be answered.

For each text, give a **Confidence Score**, which can be one of **Low**, **Medium**, **High** based on how confident you were when answering the questions in **Categories 1.1, 1.2 and 1.3**.

In case it is not clear whether a question should be answered with **Yes** or **No**, use the link provided in the link column (if available) to arrive at a more informed decision.

In case it is not clear whether a question should be answered with **Yes** or **No, and** you already checked the link, provide an answer and note a **Low Confidence Score**.

Two additional labels should be answered with **Yes** or **No**:
(1) The **Ironic Text** label is to be answered for all texts. Texts are considered ironic, if they are not meant to be taken seriously.
(2) The **Compound Claim** label is only relevant for **Category 1.1** and is defined in the respective section of the protocol.

Questions

Category 1.1: Scientific Knowledge (scientifically verifiable claims/questions)

Question: Does the text include a claim or a question that could be scientifically verified?

A claim or a question is scientifically verified if it's scientifically shown to be true OR scientifically shown to be FALSE.

Hints:

A text might contain more than a single claim or question. In that case, the text is considered scientifically verifiable if AT LEAST ONE of the claims or questions is scientifically verifiable.
Examples:

"Fauci says covid comes from France"

- Claim 1: Fauci says X -> NOT scientifically verifiable
- Claim 2: Covid comes from France-> scientifically verifiable

"Vaccines were made too fast and God disapproves vaccines"

- Claim 1: Vaccines were made too fast -> scientifically verifiable
- Claim 2: God disapproves vaccines -> NOT scientifically verifiable

For texts that contain more than a single claim or question (like in the previous examples), please answer **Yes** in the **Compound Claim** column.

A claim or question COULD be scientifically verified, IF:

1) it has probably been scientifically verified and could be found in documents created by scientists, ranging from scientific papers to official statistics from the government or from a research center, that might not have been published as a scientific paper
2) it could in theory be scientifically verified

A claim or question COULD be scientifically verified, EVEN THOUGH:

1) it seems hard to scientifically verify
2) it seems absurd
3) it is funny/ ironic and isn't intended to be taken seriously (if the text belongs to this category, answer **Yes** in the **Ironic Text** label)
4) it is not written with proper grammar/ syntax

A claim or question could NOT be scientifically verified, IF:

1) it is about an individual or the belongings of an individual (e.g indicated with a possessive 's ("Trump's family …") or a possessive pronoun ("My family …"))
2) it is about an individual's opinion or feelings
3) it is based on experience and is therefore not reproducible

Examples:

Scientifically verifiable claims or questions:

1. A scientifically verifiable claim or question that could be found in documents created by scientists
    - Cholera makes you infertile (verification could be found in scientific papers)
    - Ghana has seen a significant increase of cholera cases in 2015 (verification could be found in official statistics created by WHO scientists)
    - "Biden says vaccines are safe for children" (only taking into account the second claim "vaccines are safe", verification could be found in scientific papers)

2. A claim or question that could in theory be scientifically verified
    - "Inflation will have a huge impact on the economic stability of future generations"

3. A scientifically verifiable claim or question that is hard to verify
    - "Anyone can raise his standard of living if he's willing to work at it"

4. A scientifically verifiable claim or question that is absurd/ funny/ ironic
    - "Laughter increase life expectancy if you have live in France"

5. A scientifically verifiable claim or question that is written without proper grammar/ syntax
    - "Maxwell - parp inhibitor side effects - nausea, diarrhea and fatigue"

NOT scientifically verifiable claims or questions:

1. A claim or question that cannot be found in documents created by scientists
    - "After Poland's strict in-vitro fertilization laws, single women are losing access to their frozen embryos" (verification can be found in a law document, which isn't created by scientists)

2. A claim or question that is about an individual or an individual's belonging
    - "Obama has a higher IQ than the average US citizen."
    - "Bill Gates inserts microchips into children with vaccines"

3. A claim or question that is experiential and not reproducible
    - "The flight attendant yelled at the passengers."

Category 1.2: Reference to Scientific Knowledge

Question: Does the text include at least one reference to scientific knowledge?

Hints:
A reference to scientific knowledge could be:

1) a formal reference like a url, DOI or title of a paper/ publication/ poster/ report/ statistics/ dataset
2) an informal reference of a paper/ publication/ poster/ report/ statistics/ dataset by stating its origin (e.g. the authors, research team or university) or its name as used by the public
3) an indirect reference like a link to a news article, blog post, web site, image, etc. that includes a formal or informal reference (see (1) and (2) above)

Examples:

1) Formal reference
    - "Benchmarking State-Of-the-Art Deep Learning Software Tools https://arxiv.org/abs/1608.07249"
    - "@GregDowney1 Thought of you with this paper, "Mirroring and beyond: coupled dynamics as a generalized framework for modelling social interactions""

2) Informal reference:
    - "The study by researchers from NYU shows that sitting all day harm your health"
    - "Sitting all day harms your health as shown by this stastistic from NYU researchers"
    - "The Heinsberg study showed that COVID vaccination is important"

3) Indirect reference:
    - "Can sitting all day harm your #health? Check out these tricks to offset this via @ScienceDaily-> http://t.co/tqhgo12bO4" (Tweet includes a url to a science news article. The article includes a reference to the study 10.2215/CJN.08410814)
    - @COVID_19Watch a bit sensationalist? the abstract alone shows that this merely highlights the risks of cross species (human) transmission from the horse shoe bats of China. (Tweet includes an image of an abstract of a scientific paper)

Category 1.3: Related to scientific research in general

Question: Does the text mention a scientific context ?

Hints:
a scientific context could be:

1) scientists making a scientific discovery/ research finding
2) scientific research in general
3) scientific research efforts/ project
4) scientific research methods
5) a scientist
6) a group of publications/ posters/ reports/ statistics/ datasets
7) a journal or a conference and related entities like submission deadline and paper award

Examples:

1. Scientists making a scientific discovery/ research finding:
    - Examples:
        - In Paris in 2015, researchers discovered microplastic falling from the air. Estimated 3-10 tonnes of fibres deposited on the city each year.
        - "Scientists have found that mitochondria changes shape"
    - Counterexamples:
        - "Mitochondria changes shape"
        - "How many exe's do you have and how many do you still care about — what... you doing research about my personal life?"
        - "I just got my covid vaccine"
        - "Follow #COVID19 Appropriate Behaviour and encourage others to follow the same #Unite2FightCorona"

2. Scientific research in general:
    - Examples:
        - "Find out how social sciences & humanities research is helping us through the pandemic"
        - "Research on Covid19 Vaccines is promising"
    - Counterexamples:
        - "How many exe's do you have and how many do you still care about — what... you doing research about my personal life?"
        - "I just got my covid vaccine"
        - "Follow #COVID19 Appropriate Behaviour and encourage others to follow the same #Unite2FightCorona"

3. Scientific research efforts/ projects:
    - Examples:
        - "The new covid-19 vaccine is being thoroughly tested at the moment"
        - "#MDA hosted a FB Live event discussing the effect #COVID19 has had on research efforts in #musculardystrophy"
        - "We need more closed loop studies in people with high HbA1c and higher levels of socioeconomic deprivation."
        - "This year, we funded over 150 research grants to continue progress in finding a cure for #T1D"
        - The #Carbonfibre structure was finally delivered for the #AI4Sci projects

- Counterexamples:
    - “My new car is tested at the moment”

4. Scientific research methods:
    - Examples:
        - “After Poland's strict in-vitro fertilization laws, single women are losing access to their frozen embryos”
        - Gucci Mane and Migos are guiding me through my quantitative analysis of multivariate regressions. Just as God intended.

5. Scientist (only if the scientist is mentioned in his role as a scientist):
    - Examples:
        - “Fauci made a public statement on COVID-19”
    - Counterexamples:
        - “Fauci was born in Indiana”

6. Group of publications/ posters/ reports/ statistics/ datasets:
    - Examples:
        - Feel free to browse our publications on #malaria, #tuberculosis and #HIV / #AIDS
        - Feel free to browse our datasets on cholera cases in Westafrica from 2010-2020
        - “Posters as far as the eye can see at the #CRI2018 poster session!”
    - Counterexamples:
        - “Posters as far as the eye can see at the Comic Conference 2022”
        - “Too many books on Amazon”

7. Journal or a conference and related entities like submission deadline and paper award:
    - Examples:
        - “The #AAAI2023 takes place in June 2023”
        - Machine Learning and Applications: An International Journal (MLAIJ)
        - “Posters as far as the eye can see at the #CRI2018 poster session!”
        - “Best paper award for an outstanding paper published in the Journal of AI Research in the last 5 years.
    - Counterexamples:
        - “I bet harry writes poetry in that journal of his along with songs and inspiration quotes to self”
        - “Posters as far as the eye can see at the Comic Conference 2022”
        - “Philippines Pre-order of Kim Jaejoong's 2nd Single DEFIANCE -- Order form submission deadline is until September 22 only.”

Category 2: NOT science-related

The text belongs to neither **Category 1.1**, **Category 1.2** nor **Category 1.3**.