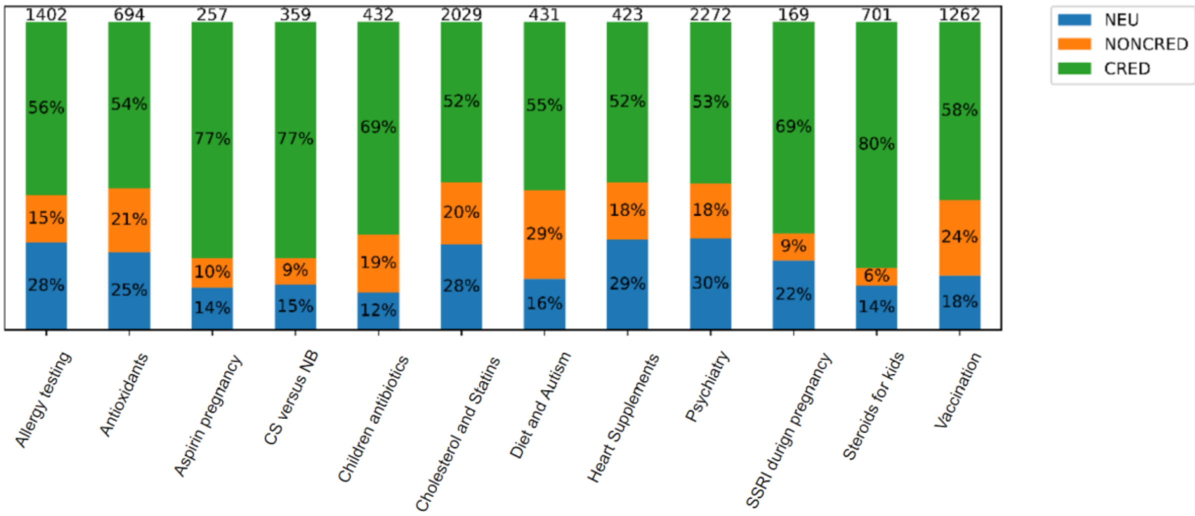
Fighting medical disinformation is an increasingly important problem. As of today, automatic systems for assessing the credibility of medical information do not offer sufficient precision to be used without human supervision, and the involvement of medical expert annotators is required. That's why I decided to design a tool for credibility annotation of medical online texts that would incorporate algorithms as well as human expert effort. However, during my research I encountered two important problems:

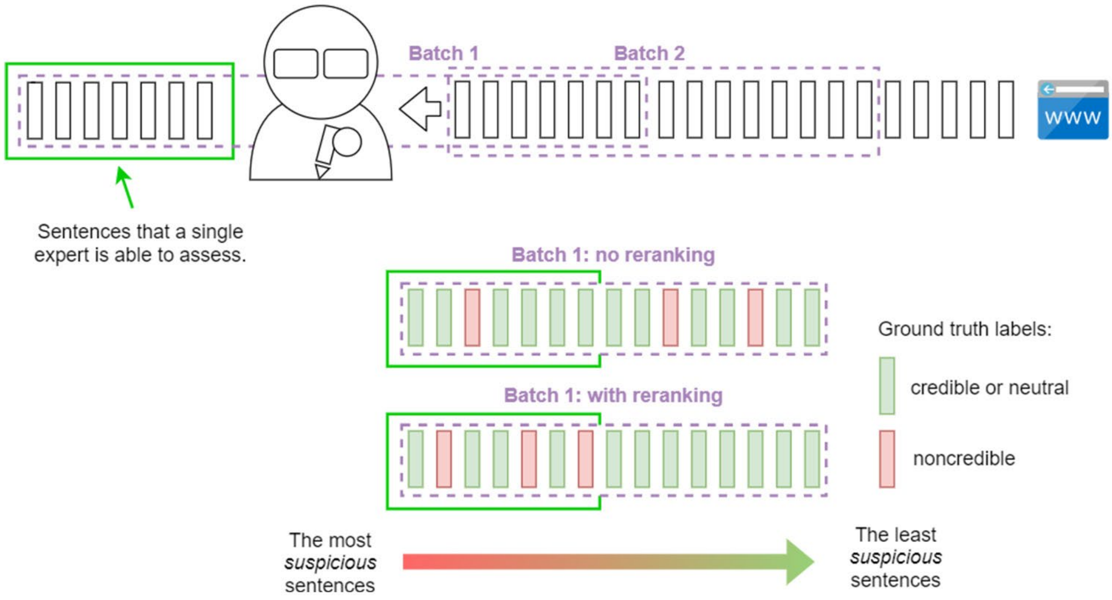
The first one: medical misinformation is sparse. Why would an expert, whose time is so precious, spend hours to review many credible samples of texts? See the figure 1 to get an idea. Those are the percentages for the dataset of texts that were already chosen to cover controversial topics (as stated by medical experts themselves).



Ok, but if misinformation in Online Health Information (OHI) is so sparse, why would we care at all? Because sparcity does not mean that misinformation have little impact!

The second problem: even experienced medical professionals find it challenging to assess the truthfulness of online medical information. What is considered to be "true" in the domain of medicine is often subject to a very complex context. This context is provided by external medical knowledge and clinical practice. Medical professionals often focus on the possible impact of health information on the choices made by patients rather than evaluate the factual correctness of a statement. In other words, a factually correct statement may still inflict health damage on patients when presented mischievously or in isolation. The phrase "For starters, statin drugs deplete your body of coenzyme Q10 (CoQ10), which is beneficial to heart health and muscle function", despite factual correctness, would raise objections from medical professionals as it may discourage a patient from taking statins. In this example, the expert uses external knowledge from their clinical practice that for patients requiring statin therapy, its benefits far outweigh the potential risks associated with coenzyme Q10 deficiency. This additional context of online health information evaluation makes it extremely difficult to frame the task in terms of machine learning.

To address the first problem I have presented a framework for the optimization of the utilization of medical experts' time when evaluating the credibility of OHI. The general idea is sketched in the second figure:



Considering the constant stream of texts available through the World Wide Web (top right part of the picture) and the limited batch of sentences that an expert is able to assess (top left), we would like to focus on the most "suspicious" content. One way to achieve that is filtering out credible sentences (or put them at the end of the annotation queue). Thus, me and my colleagues had trained classifiers that can filter out credible and neutral medical claims with very high precision exceeding 90% for most medical topics considered in our study (vaccination, allergy testing, children antibiotics, steroids for kids, antioxidants, cholesterol & statins, and C-section vs. natural birth). The potential human-in-the-loop factchecking system that our solution provides may increase the probability that a medical expert will encounter a non-credible medical statement in the annotation batch by the factor of 2.

To address the second problem, we performed the qualitative analysis of the annotated samples of texts. **The aim of this analysis was to curate the annotation protocol for the future annotations**. We selected sentences that contain similar words and statements but differ in the narrative details that skewed the experts' judgments. We have identified 4 types of false and misleading narratives that occur frequently in the "non-credible" class. These narratives are as follows:

Slippery slope: The sentence is factually true, but the consequences of the presented fact are exaggerated.

Example:

Hence, while the drug might synergise with a statin to prevent a non-fatal (or minor) heart attack, it seems to increase the risk of some other equally life-threatening pathology, resulting in death.

Cholesterol also helps in the formation of your memories and is vital for neurological function.

Hedging: The sentence is factually incorrect, but there is a part of it that softens the overtone of the presented statement.

Example:

However, cholesterol content should be less of a concern than fat content.
[CRED]

Coenzyme Q10 supplements may help prevent statin side effects in some people, though more studies are needed to determine any benefits of taking it.
[CRED]

The FDA warns on statin labels that some people have developed memory loss or confusion while taking statins.
[CRED]

Alleged negative consequences: The sentence is mostly factually true, but given the context of the expert's experience, there is a risk that the presented information may lead the patient to act contrary to current medical guidelines.

Examples:

For starters, statin drugs deplete your body of coenzyme Q10 (CoQ10), which is beneficial to heart health and muscle function.

Cholesterol is a waxy, fatty steroid that your body needs for things like: cell production.

Twisting words: the presence of a single word changes the overtone of the sentence.

Examples:

Statins may slightly increase the risk for Type 2 diabetes, a condition that can lead to heart disease or stroke. [CRED]

For example, it may be enough to eat a nutritious diet, exercise regularly, and avoid smoking tobacco products. [NONCRED]
versus

Eating a healthy diet and doing regular exercise can help lower the level of cholesterol in your blood.
[CRED]

Having defined and presented to the annotators such classes of misleading narratives, we are now able to increase the, so called, inter-expert agreement rate.

Although there are still many problems and challenges arising regarding the topic of OHI credibility tagging, some important work have been done and some milestones have been reached. I hope you enjoyed this quick summary!