# Expert Insights into Studying Misinformation with Natural Language Processing Methods

## An Interview with Arkaitz Zubiaga

*Not only in times of crisis and unrest is the internet and particularly social media being used as a critical source of information for the largest parts of our societies. However, in order for the "social web" to function as such an immediate and trustworthy information ecosystem, large scale and largely automated (computational) methods are required to mitigate the impacts of wrong or misleading information that is purposefully being created and shared online.*

*We talked to Arkaitz Zubiaga about some of the established practices in dealing with misinformation as well as novel challenges arising with the proliferation of large language models. Arkaitz is an Associate Professor and Co-leader of the Social Data Science Lab at the Queen Mary University of London. His work on the development and improvement of natural language processing (NLP) methods for the analysis of online and social media data helps to understand and tackle some of the problematic aspects of online communications, including misinformation. In the interview, Arkaitz talks us through the different steps of his pipeline for detecting misinformation, the roles of experts and journalists in it, and disentangles some terminological confusion between "rumors" and "fake news".*

*Indira Sen and Leon Fröhling conducted the interview with Arkaitz on June 5, 2023, during the International Conference on Web and Social Media (ICWSM-23) in Limassol, Cyprus. The transcript was edited for clarity and length.*

**GESIS: Thank you, Arkaitz, for giving us this interview. What are you working on at the moment – especially in the context of online misinformation and related areas?**

**Arkaitz Zubiaga:** Within the topic of misinformation, I am working primarily on automated fact-checking. This involves a number of problems starting with picking up the information

that we need to verify for giving a verdict. This includes several steps: Asking how we actually go from taking a text, to identifying some evidence that will help us verify that piece of text, to then putting together the original claim or text with the evidence, and lastly to give a verdict. This involves research in both natural language processing (NLP) and computational social science. We use many different techniques for classifying text and for understanding whether a claim is correct or not.

**GESIS: Could you tell us a bit about how you entered this area of research and what motivated you to study it?**

**Arkaitz Zubiaga:** When I did my PhD, I was already doing NLP for social media. After these things evolved towards working with journalists on a European project [1], where we were trying to verify social media rumors as they were circulating, and to gather information about the rumors to try and verify them. More recently, my work evolved from focusing on rumors to working with fact-checking organizations and identifying claims of journalistic interest and verifying them.

**GESIS: You already indicated that the field has changed quite a lot since you started. Has the whole information online topic taken off in recent years?**

**Arkaitz Zubiaga:** I would say it started very vaguely around a decade ago with very preliminary research. At that moment, there were not so many advanced methods as we have now. There was some initial research trying to understand how fake information circulated, why people were circulating fake information, but it was rather general or introductory. And now this has evolved into looking at lots of different types of misinformation, and into building different methods.

**GESIS: In the train of that development, did some of your internal understandings and perceptions of the problem change? Did you have any major reflection point where you would say "I understood something in a new way and that helped me tackle this from a different perspective?"**

**Arkaitz Zubiaga:** Things have evolved definitely because I have learned a lot within this process. One of the things that changed a lot is that I used to work much more on identifying which pieces of information to verify rather than verifying them.

We chose identifying rumors and see what is circulating and what we might want to verify. By then I realized that we could actually make this bigger and go further and build methods that could go from having those claims identified to linking them to pieces of evidence and verifying them. This was not obvious a decade ago when research was starting. There was not such a well-defined pipeline for doing research in automated fact-checking and misinformation verification.

**GESIS: Getting a bit deeper into misinformation and especially in the computational social science and NLP context, could you tell us why it is such an important topic and where the computational aspects can help?**

> *Experts assessing the veracity of claims and a handful of fact-checkers are not enough for the scale of this problem.*

**Arkaitz Zubiaga:** It definitely is something that has a huge impact in society. Having a lot of fake information circulating can cause severe damages in society. It can have an impact on people's mental health but also on the actions that people take, e.g., how they vote, or how they behave on a daily basis based on what they perceive or understand that is actually happening in society. It is because of this impact on actual behavior that it is so crucial to have an accurate understanding of what is actually happening.

> *In those parts where you can scale up the identification of claims, the finding of evidence, identifying what is becoming viral, computational methods can help a lot.*

Experts assessing the veracity of claims and a handful of fact-checkers are not enough for the scale of this problem. Having computational tools can help scale processes by automatically identifying the claims that need to be fact-checked. Questioning what is circulating, what is becoming viral and popular should be a priority to fact-check to identifying the pieces of evidence that can be linked to the claims. I think that the verification part itself is more difficult to tell for an algorithm, or: it is more difficult for an expert to trust automation at this stage. In those parts where you can scale up the identification of claims, the finding of evidence, identifying what is becoming viral, computational methods can help a lot.

**GESIS: What are the current challenges and developments in misinformation research? Looking at things like ChatGPT where we see the first papers trying to fact-check and assess the veracity of information using these large language models – is that something that you would also like to try in your research?**

**Arkaitz Zubiaga:** There are two possible angles on how ChatGPT can impact misinformation research. One is that ChatGPT can be harmful by generating lots of misinformation very cheaply and generating huge scales of misinformation that can circulate and make even more damage. In that sense it is definitely negative and really harmful. On the other hand, on a more positive side, we could hope, perhaps, that ChatGPT can help generate more samples of misinformation so that the models can learn what misinformation can look like on a larger scale.

I am a bit more reluctant with this direction personally because I think it is difficult to have a ChatGPT generate realistic samples of misinformation that can resemble what humans

would generate or what can circulate in realistic settings. I am not saying it is impossible, but I think it is difficult at this point in time. So, I am not very convinced so far how ChatGPT could – positively – impact misinformation and misinformation research. I think, for the time being, it can cause more damage than benefits.

**GESIS: Related to that, do you encounter any prevalent misconceptions or misunderstandings in your area of research?**

> *"Fake" [...] is not the same as "unverified".*

> *I think calling misinformation "fake news" is not an appropriate term because there are many other terms that we can use for more specific things.*

**Arkaitz Zubiaga:** There are quite a few, especially regarding the terminology used and how the concepts are employed. I have seen confusions for example with the term "rumors" which I have always used to refer to unverified information that is circulating, something that is being talked about but there is not enough evidence or background information to verify with reputable sources.

From my point of view rumors are information that are circulating as unverified, and they could turn out to be true or false when more evidence emerges, or it could remain unverified forever for lack of evidence. I have seen that sometimes "rumors" are referred to as "fake information" or as being "fake", which is not the same as "unverified".

Another one that I do not quite like is the "fake news" term because it is such a broad term that has been used in many contexts. It is difficult to define "fake news". It is not that bad to use "fake news" for fabricated news articles or something similar.  If, for example, ChatGPT generates news articles which are inaccurate, we could refer to it as fake news because they are "fakely" generated.  I think calling misinformation "fake news" is not an appropriate term because there are many other terms that we can use for more specific things.

**GESIS: Connected to the topic of rumors being unverified claims that could turn out to be both true or false: how do you think about the role of experts in this whole misinformation research compared with other approaches? For example, on Twitter you have community notes (previously called 'birdwatch'), where you would use the wisdom of the crowd to assess whether something is true or false. Do you think that this role should be reserved for experts, or would you say that something like those community notes can also play a beneficial role?**

**Arkaitz Zubiaga:** I think both can be beneficial. Experts are trained to do this, and they are more qualified to help out with this. We should rely especially on them but that does not mean that we cannot also benefit from the other less trained people who try to do a good job. I think we can definitely benefit from the community, but we have to be more careful in seeing who is giving us an opinion here. Why would this person be trustworthy is something to check first. But it is mainly experts who can provide datasets, knowledge, and expertise to help develop the models and train the community.

**GESIS: For the datasets that you are using, are you mainly cooperating with experts, fact checkers, or journalists?**

**Arkaitz Zubiaga:** In the past I worked with journalists collecting datasets that can support this misinformation analysis. One of the challenges is that we often rely on publicly available datasets from fact checkers, and one of the problems with those datasets is that they are not representative of the state of misinformation because they take whatever the fact checkers report on their website. It is a selection of what they believe are the main things that they should be reporting about. They tend to be more skewed towards fake and misinformation than accurate information because they prioritize reporting about this to society. That carries the risk that the dataset is not really representative of all types of information that we receive and that we may see in a realistic scenario.

**GESIS: Building on that and for practical matters like building pipelines for rumor verification and rumor detection – these are also rather sensitive tasks or can be for certain topics –, how would you set up such types of automated pipelines, and how do you think about mixing human evaluation or human oversight with computational pipelines?**

> *In this project one of the main purposes is to ask how to combine knowledge input from humans with automated methods. In a few years, I can give you a better response to that.*

**Arkaitz Zubiaga:** That is a big and open question, I am not sure we have solutions to just go and merge the two outputs. We are currently working on it in a European project, called "hybrids" [2]. In this project, one of the main purposes is to ask how to combine knowledge input from humans with automated methods. In a few years, I can give you a better response to that. I think that we need to combine them both and that we need to take the knowledge from both – because they are complementary.

**GESIS: Connected to what you said before, you create and share a lot of datasets on rumors and misinformation, and you have to put that out in the community. This question of what datasets to use for research is a very important one and maybe you could**

**give us some insight on how you approach the topic. How do you decide which dataset to use for you your research, are there any criteria that you look at for assessing whether it suits your purposes or not?**

**Arkaitz Zubiaga:** There are a few things that you need to look at. One of them is whether the data is representative of something meaningful. That is always difficult, so it is good to know which data sampling strategy has been followed in creating the dataset. The other one is the annotation quality – who did the annotations, how did they make the final judgment, how did they combine opinions with different experts? Then there are other secondary factors like the size of the dataset and how the dataset is published – if it is fully available with all required meta data etc.

**GESIS: Are there any types of tutorials, guidelines or packages that you would suggest? You did a workshop [3] on this topic, which is a good introduction to the research area, but other than that, do you have further suggestions?**

**Arkaitz Zubiaga:** There are quite a few good handbooks from fact checkers and journalists which are very useful to read and use in combination with reading about NLP, computational social science and computational tools in general. I think it is really important to be interdisciplinary and read about what journalists and fact checkers are saying. There are some good ones from FullFact [4] who have written guidelines of how they do fact checking from First Draft [5]. There is one from Craig Silverman on verification [6]. Those are really nice handbooks to understand the problem from a more practical point of view, from a journalistic angle.

**GESIS: If you could make a wish to the universe for something that helps you solve all your misinformation problems and troubles – like a package, an app, a theory, or a whole research agenda that you could pursue, is there anything that you would really wish for?**

**Arkaitz Zubiaga:** One is to be more mindful about sharing misinformation and about the impact that this can have, so: do not contribute to that. And the other one would be for people to ask people to volunteer to help out with this, to collaboratively collect datasets by sending reports of misinformation they find and try and build something larger scale between a lot of us.

**GESIS: Thank you, Arkaitz, for this interview!**

## References

1　Kochkina, E., Liakata, M., & Zubiaga, A. (2018). All-in-one: Multi-task Learning for Rumour Verification. *ArXiv*. https://doi.org/10.48550/ARXIV.1806.03713

2　Hybrids (s.d.). https://hybridsproject.eu/jobs/dc8/ [retrieved Nov 1, 2023]

3　Mediate (June 5, 2023). Workshop at the 17th International AAAI Conference on Web and Social Media (ICWSM). https://digitalmediasig.github.io/Mediate2023/ [retrieved Nov 1, 2023]

4　Full Fact (s.d.). How we fact check. https://fullfact.org/about/how-we-fact-check/ [retrieved Nov 1, 2023]

5　First Draft (2015-2022). [Fact checking archives.] https://firstdraftnews.org/articles/tag/fact-checking/ [retrieved Nov 1, 2023]

6　Silverman, C. (Ed.) (2014). *Verification handbook. An ultimate guideline on digital age sourcing for emergency coverage*. European Journalism Centre. https://verificationhandbook.com/ [retrieved Nov 1, 2023]