

Ph.D. research topic

- Title of the proposed topic: Natural language counter-argumentation to fight online disinformation
 - Research axis of the 3iA:
 - Axes 2: AI FOR INTEGRATIVE COMPUTATIONAL MEDICINE
 - Axes 1: CORE ELEMENTS OF AI
 - **Supervisor (name, affiliation, email): Serena Villata (Université Côte d'Azur, CNRS, Inria, I3S), email: serena.villata@univ-cotedazur.fr**
 - Potential co-supervisor (name, affiliation): Elena Cabrio (Université Côte d'Azur, CNRS, Inria, I3S), email: elena.cabrio@univ-cotedazur.fr
 - The laboratory and/or research group: WIMMICS (<http://wimmics.inria.fr/>) is a research team of Université Côte d'Azur (UCA), Inria, CNRS. The research fields of the team are graph-oriented knowledge representation, reasoning and operationalization to model and support actors, actions and interactions in web-based epistemic communities.
-

Apply by sending an email directly to the supervisor.

The application will include:

- **Letter of recommendation of the supervisor indicated above**
 - Curriculum vitæ.
 - Motivation Letter.
 - Academic transcripts of a master's degree(s) or equivalent.
 - At least, one letter of recommendation.
 - Internship report, if possible.
-

- Description of the topic:

Online social media platforms try to limit the virality of disinformation spread for example through content moderation. While these measures show some kind of effectiveness in limiting the diffusion of misleading or fake information, one big issue still remains unsolved: identifying disinformation and reporting its status to the users is not enough to counter it [2].

The **goal** of the PhD program is to address that urgent need by designing intelligent (semi-)automated ways to generate counter-arguments to fight the spread of disinformation online [1]. The idea is to avoid the undesired effects that come with content moderation, such as overblocking, when dealing with disinformation online, and to directly intervene in the discussion (e.g., Twitter threads) with textual arguments that are meant to counter the fake

content as soon as possible, and prevent it from further spreading. The idea is to obtain high quality counter-arguments while reducing efforts and supporting human fact-checkers in their everyday activities.

The **main objectives** of the PhD program therefore are:

1. Definition of a general model of counter-argument production in real life. The use case will be Covid-19 data for which verified sources of information exist as well as fact-bounded arguments.
2. Dataset creation, by testing several hybrid strategies to collect counter-arguments against disinformation, i.e., nichesourcing (collecting counter-arguments in a controlled setting, where these arguments are written by fact-checkers, i.e., experts in counter-arguments production), crowd-sourcing (writing possible responses to a set of annotated messages containing fake information), and synthetic counter-argument generation obtained by fine-tuning deep neural architectures.
3. Generation of natural language counter-arguments, starting incrementally from the generation of claims countering the hate content towards the generation of full argumentative structures composed by evidence and the supported claim.
4. Evaluation of the quality of the generated counter-arguments with respect to the lexical and semantic diversity, but also with respect to other quality dimensions like cogency, effectiveness, reasonableness dimensions.
5. Definition of a task-oriented dialogue system which provides the generated counter-arguments about a piece of disinformation and engage with the user in critical interactive dialogues.

References

- [1] Elena Cabrio, Serena Villata. Five Years of Argument Mining: a Data-driven Analysis. Proceedings of 27th International Joint Conference on Artificial Intelligence (IJCAI 2018), pages 5427-5433.
- [2] Jacky Visser, John Lawrence, Chris Reed . Reason-Checking Fake News. Communications of the ACM, November 2020, Vol. 63 No. 11, Pages 38-40.
- [3] CNPEN bulletin "Ethical issues related to online disinformation and misinformation spreading" - co-rapporteurs E. Didier, S. Villata, C. Zolynski

Keywords:

Natural Language Processing, Disinformation, Fake News, Argument Mining, Counter-argumentation, Natural Language Generation

Skills and profile:

- Master degree in Data Science, Computer Science or Computational Linguistics is required.
- Programming skills are required.
- Knowledge of Natural Language Processing and Machine Learning is preferred.
- Fluent English required, both oral and written. French is appreciated but not mandatory.