

**Titre de la thèse/Thesis title :** Explainable and Multi-Modal DL/ML Models for Extreme Narrative Detection in the Online Social Discourse.

**Laboratoire d'accueil / Host Laboratory :** ETIS (MIDI team) (<https://www.etis-lab.fr/midi/>)  
AGORA (<https://cyagora.cyu.fr/>)

**Spécialité du doctorat préparé/ Ph.D degree in :** Computer Science (Informatique)

**Mots-clefs / Keywords :** Social Unacceptable Discourse, Extreme Narrative, Language Model, Deep Learning, Machine Learning, Multi-Modal ML, Knowledge Base.

**Descriptif détaillé de la thèse / Job description :**

Increased polarization triggered by social protest movements, the Covid-19 crisis and the war in Ukraine are historical events that have recently favored extremist narratives in public and online debates. **Extremist (a.k.a Extreme) narratives (EN)** constitute counter-narratives in the sense that they challenge mainstream worldviews and social interpretations of major events in many kinds of public debate e.g., social media, parliamentary interventions, journals, books, and many others. This thesis offer is integrated in the context of the Horizon Europe ARENAS project (Grant agreement ID: 101094731), coordinated by CY Cergy Paris Université, and aims to contribute to Work Package 2 dedicated to the definition, identification and detection of extremist narratives. (<https://cordis.europa.eu/project/id/101094731>).

**In this Ph.D. thesis**, we want to study the EN characterization, modeling and automatic detection. Specifically, we note that the extremist narratives analysis should not only be seen from a radicalization/terrorism viewpoint, for which a rich Machine Learning (ML) literature already proposes multiple solutions [1,2,3]. We observe that EN must be studied in a more general context that concerns different kind of values such as people democracy, citizenship, rights, etc., which do not necessarily assume a violent or hatred sentiments. EN modeling cannot be only isolated to violent and extreme language features [4,5,6,7,8,9,10], but it must also consider a wider spectrum of narrative elements such as the beliefs, traits, practices of a collectivity, etc. that identifies a group of people sharing the same identity [14,15].

**The principal thesis objective** is to propose new (DL/ML) tools that characterize extremist narratives in corpora from different contexts (social media, political debates, transcripts, etc.).

We argue that EN modeling choice is not only restricted to text but must effectively consider other types of data, i.e., Graphs, Images, and Knowledge Base. In this case, we want to focus on multi-modal knowledge extraction, which is a challenging topic in Machine Learning.

The existing multi-view machine learning approaches [22, 24, 26] usually are not adapted for multi-modal data [21, 23, 25] or use the same similarity/distance measure for all the views.

A crucial objective of our research is to propose novel multi-modal knowledge extraction methods to detect extremism narratives and characterize them.

**The successful candidate** will work in close collaboration with language experts (from Heinrich Heine University of Düsseldorf and from Institute of Contemporary History-Ljubjana University) that will provide linguistic expertise and validation, along with labelled corpora from heterogeneous online (multi-modal) content. Interactions will be made with work already in progress at Cergy on forensic linguistics, the analysis of fake news, and digital discourse in a political context [16, 17,18, 19].

**Références bibliographiques / Bibliography**

[1] Paula Fortuna and Sérgio Nunes. A survey on automatic detection of hate speech in text. ACM Comput. Surv., 51(4):85:1–85:30, 2018.

- [2] Mariam Nouh, R.C. Nurse, and Michael Goldsmith. Understanding the radical mind: Identifying signals to detect extremist content on twitter.
- [3] Anna Schmidt and Michael Wiegand. A survey on hate speech detection using natural language processing. In SocialNLP@EACL 2017.
- [4] Segun Taofeek Aroyehun and Alexander Gelbukh. Aggression detection in social media: Using deep neural networks, data augmentation, and pseudo labeling. In Proceedings of the First Workshop on Trolling, Aggression and Cyberbullying (TRAC-2018), 2018.
- [5] Jing Qian et al. A benchmark dataset for learning to intervene in online hate speech, 2019.
- [6] Lara Grimminger and Roman Klingler. Hate towards the political opponent: A Twitter corpus study of the 2020 US elections on the basis of offensive speech and stance detection. In Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, April 2021.
- [7] Sean MacAvaney, Hao-Ren Yao, Eugene Yang, Katina Russell, Nazli Goharian, and Ophir Frieder. Hate speech detection: Challenges and solutions. PLOS ONE, 2019.
- [8] et al. Marcos Zampieri. Predicting the type and target of offensive posts in social media, 2019.
- [9] Sayar Ghosh Roy, Ujwal Narayan, Tathagata Raha, Zubair Abid, and Vasudeva Varma. Leveraging multilingual transformers for hate speech detection, 2021.
- [10] Steve Durairaj Swamy, Anupam Jamatia, and Bjorn Gambäck. Studying generalisability across abusive language detection datasets. In Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL), 2019.
- [11] Betty van Aken, Julian Risch, Ralf Krestel, and Alexander L'oser. Challenges for toxic comment classification: An in-depth error analysis, 2018.
- [12] Bin Wang, Yunxia Ding, Shengyan Liu, and Xiaobing Zhou. Ynu wb at hasoc 2019: Ordered neurons lstm with attention for identifying hate speech and offensive language. In Fire, 2019.
- [13] Lanqin Yuan and Marian-Andrei Rizoiiu. Detect hate speech in unseen domains using multi-task learning: A case study of political public figures, 2022.
- [14] Berger, J. M. (2018). Extremism. The MIT Press. <https://doi.org/10.7551/mitpress/11688.001.0001>
- [15] Glazzard, Andrew. *Losing the Plot: Narrative, Counter-Narrative and Violent Extremism*. International Centre for Counter-Terrorism, 2017. JSTOR, <http://www.jstor.org/stable/resrep29416>.
- [16] Longhi, Julien. "Using digital humanities and linguistics to help with terrorism investigations." *Forensic science international* 318 (2021): 110564
- [17] Degeneve, Clara, Julien Longhi, and Quentin Rossy. "Analysing the digital transformation of the market for fake documents using a computational linguistic approach." *Forensic Science International: Synergy* 5 (2022): 100287.
- [18] Longhi, Julien. "Linguistic Approaches to the Analysis of Online Terrorist Threats." *Language as Evidence: Doing Forensic Linguistics*. Cham: Springer International Publishing, 2022. 439-459.
- [19] Longhi, Julien. "Mapping information and identifying disinformation based on digital humanities methods: From accuracy to plasticity." *Digital Scholarship in the Humanities* 36.4 (2021): 980-998.
- [20] Grozavu N., Y. Bennani, B. Matei, K. Benlamine. Multi-view clustering based on non-negative matrix factorization. In W. Pedrycz and S.-M. Chen, editors, *Recent Advancements in Multi-View Data Analytics, Lecture Notes in Computer Science*. Springer, 2022
- [21] Grozavu N., Khalafaoui Y., Matei B., Goix L.-W., "Multi-modal Multi-view Clustering based on Non-negative Matrix Factorization". [SSCI 2022](https://doi.org/10.1007/978-3-030-61111-1_13): 1386-1391
- [22] Yang, Y. and Wang, H., 2018. Multi-view Clustering: A survey. *Big Data Mining and Analytics*, 1(2), pp.83-107.
- [23] Hu, D., Nie, F. and Li, X., 2019. Deep Multimodal Clustering for Unsupervised Audiovisual Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (9248-9257).
- [24] Devlin, J., Chang, M.W., Lee, K. and Toutanova, K., 2018. Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv preprint arXiv:1810.04805*.
- [25] P. P. Liang, Y. Lyu, X. Fan, Z. Wu, Y. Cheng, J. Wu, L. Chen, P. Wu, M. Lee, Y. Zhu, R. Salakhutdinov, L.-P. Morency. MultiBench: Multiscale Benchmarks for Multimodal Representation Learning. *NeurIPS 2021 Datasets and Benchmarks Track*
- [26] Sublime J., Matei B., Cabanes G., Grozavu N., Bennani Y., Cornuejols A. (2017), "Entropy Based Probabilistic Collaborative Clustering", *Pattern Recognition*, 2017, *The Journal of Pattern Recognition Society*, Elsevier

**Important :**

Applicants should contact via email Michele Linardi ([michele.linardi@cyu.fr](mailto:michele.linardi@cyu.fr)), Julien Longhi ([julien.longhi@cyu.fr](mailto:julien.longhi@cyu.fr)) and Nistor Grozavu ([nistor.grozavu@cyu.fr](mailto:nistor.grozavu@cyu.fr)) with:

- A full curriculum vitae, including a summary of previous research experience.
- A transcript of higher education records.
- A one-page research statement discussing how the candidate's background fits the proposed topic.
- Two support letters of persons that have worked with them.

**The deadline of the application is: June 4th, 2023 (11h59 pm AoE).**

**Financement :**

Début du contrat : Octobre, 1, 2023

Salaire mensuel brut : ~1975€ (Des missions doctorales sont possibles + ~400 euros brut)

***This Ph.D Thesis is founded by the European Union's Horizon Europe research and innovation program under grant agreement No 101069740.***

**Direction de la thèse:/ Thesis Supervisor:**

PR Julien Longhi (<https://www.cyu.fr/m-julien-longhi>) [julien.longhi@cyu.fr](mailto:julien.longhi@cyu.fr)

PR Nistor Grozavu ( <https://grozavu.fr> ) [nistor.grozavu@cyu.fr](mailto:nistor.grozavu@cyu.fr)

**Encadrement de la thèse : co-directeur(s) et co-encadrant(s)**

Michele Linardi MCF (<https://www.etis-lab.fr/2022/01/12/michele-linardi/>) [michele.linardi@cyu.fr](mailto:michele.linardi@cyu.fr)

**Profil demandé / Preferred Qualification**

The candidate must fit the following requirements:

- Master's degree in computer science or data science.
- Advanced programming skills in Python (C++/Java is a plus).
- Strong mathematical background, including Linear Algebra and Statistics.
- Research experience in Machine learning, Deep Learning and Data Mining.
- Fluency in written and spoken English is essential.