CounteR

NEWSLETTER

DECEMBER 2023 | ISSUE #8

Editorial

Dear CounteR Community,

The project team is thrilled to announce the start of an exciting phase of the CounteR Project's life: as we venture into the final stretch, the upcoming four months are shaping up to be intense yet incredibly rewarding. This journey began with the **launch of the CounteR Platform's third pilot** in mid-December 2023, setting the stage for what promises to be the culmination of hard work, dedication, and collaboration among our esteemed partners.

We have lined up a number of events that would lead CounteR to achieving its planned project milestones. Here is a glimpse of what lies ahead:

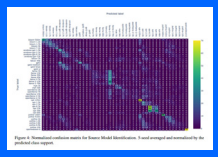
- **Second CounteR Webinar**, tentatively scheduled for 25 January 2024: AST, IMG, INS, ELTE and EI experts will discuss the role of large language models in countering radicalisation – this is the heart of CounteR's added value. Stay tuned to the programme's final version and mark your calendars for subscribing to attend. Details will come soon through our website and our accounts on X and LinkedIn. For reference, check the outcomes of our first webinar.

- **Sixth CounteR Technical Meeting & Fourth CounteR Platform Piloting Event**, planned for 6-8 February 2024 in Lisbon. Hosted by PJ, this event is set to foster insightful exchanges by consortium partners as we drive CounteR closer to fulfilling its final objectives.

- **CounteR Project Spring Conference:** hosted by UCSC in Milan on 14-15 March 2024, this event is a key moment of our journey. The conference will bring together experts and practitioners to present project results and discuss experiences from the field, related to TRA-Is and radicalisation.

- **External Stakeholders Demo Event**: as March draws to a close, we will be hosting a demonstration event for the CounteR Platform, targeted to external stakeholders. This event will reinforce the progress that the project has achieved – it will be an opportunity to present our accomplishments to a wider audience of potential users.

ISSUE #8 HIGHLIGHTS



• The Newest CounteR Scientific Publications: A Digest

(p. 3)

• CounteR attended DG HOME's Annual CERIS Event in Brussels

(p. 2)

- **Joint Activities with the TCO Cluster** – We are proud to announce that CounteR has entered into collaboration with the Terrorist Content Online (TCO) Cluster. Three ISF-funded projects – ALLIES, FRISCO, and TATE – jointly run this initiative to raise awareness about the new regulation on terrorist content online. Plenty of opportunities for synergies and shared visions with the CounteR team!

These events will bring CounteR closer to achieving the project goals and objectives and will be extensively covered in the next two (and last) editions of the CounteR newsletter. All our readers are most welcome to participate, attend, join, and interact: stay tuned to the updates as they become available.

As we navigate through this critical phase, let us continue leveraging our strengths and driving innovation within the Counter community for a lasting impact within our industry. The next months will be the culmination of our collective efforts and, for all of us, 2024 will be full of anticipation and promise.

Warmest wishes for a happy and prosperous New Year!

Sincerely yours,
Dr. Borislav Mavrov,
CounteR Dissemination Manager
European Institute Foundation (EI)

COUNTER ATTENDED DG HOME'S ANNUAL CERIS EVENT IN BRUSSELS

On 14-15 December 2023, the CounteR project, represented by EI, participated in the annual event of CERIS – the Community for European Research and Innovation for Security.

The CERIS conference, entitled "*Fighting Crime and Terrorism/Resilient Infrastructure*", was organised by the European Commission's DG Home Affairs with the objective to bring together security research practitioners, policy makers, researchers, civil society and industry to discuss cross-cutting topics that have a broad and horizontal impact on research and innovation in this domain.

Among the participants and panellists were the European Forum for Urban Security, the European Citizen Action Service, Fondazione Safe, Public Safety Communication Europe, and others, including projects and consortia funded through Horizon Europe and the Internal Security Fund.

"The CERIS event was a great opportunity to meet with colleagues from across the research and security ecosystem. A key takeaway that emerged from the event is that CSOs should serve as a bridge between academics, citizens and LEAs", Apostol Apostolov from EI, representing the CounteR consortium, commented.

More information about the CERIS event is available in the CounteR website publication.

THE NEWEST COUNTER SCIENTIFIC PUBLICATIONS: A DIGEST

A collection of newly-released scientific publications, authored by researchers and experts from the CounteR consortium, has been made available on CounteR website's Publications Section***. In Issue #8, the newsletter dedicates space to presenting a digest of these specialised articles.***

ITADATA Conference – Project Presentation and Audience Engagement | CINI

Published on 12 October 2022

Authors: Marco Anisetti, Michelangelo Ceci

A scientific publication was presented to the audience of the ITADATA2022 conference on big data and data science, conducted in September 2022 in Milan – Italy. The paper was presented by Marco Anisetti (Università degli Studi di Milano) and Michelangelo Ceci (Università degli Studi di Bari) during the Workshop on active projects of the CINI lab on big data.

SAIRUS: Spatially-aware identification of risky users in social networks | CINI

Published on 29 May 2023

Author: Antonio Pellicani, Gianvito Pio, Domenico Redavid, Michelangelo Ceci

This paper contributes towards an automated identification of risky users in social networks. Specifically, we propose a novel system, called SAIRUS, that solves node classification tasks in social networks by exploiting and combining the information conveyed by three different perspectives: the semantics of the textual content generated by users, the network of user relationships, and the users' spatial closeness, derived from the geo-tagging data associated with the posted contents.

Analysing Zero-Shot Transfer Scenarios across Spanish variants for Hate Speech Detection | INRIA

Published on 6 June 2023

Author: Galo Castillo-lópez, Arij Riabi, Djamé Seddah

In this work, we present the results of a thorough analysis of hate speech detection models performance on different variants of Spanish, including a new hate speech toward immigrants in Twitter data sets. Using mBERT and Beto, a monolingual Spanish Bert-based language model, as the basis of our transfer learning architecture, our results indicate that hate speech detection models for a given Spanish variant are affected when different variations of such language are not considered. Hate speech expressions could vary from region to region where the same language is spoken. Our new dataset, models and guidelines are freely available.

Enriching the NArabizi Treebank: A Multifaceted Approach to Supporting an Under-Resourced Language | INRIA

Published on 10 June 2023

Author: Arij Riabi, Menel Mahamdi, Djamé Seddah

In this paper, we address the scarcity of annotated data for NArabizi, a Romanized form of North African Arabic used mostly on social media, which poses challenges for Natural Language Processing (NLP). We introduce an enriched version of NArabizi Treebank (Seddah et al., 2020) with three main contributions: the addition of two novel annotation layers (named entity recognition and offensive language detection) and a re-annotation of the tokenization, morpho-syntactic and syntactic layers that ensure annotation consistency.

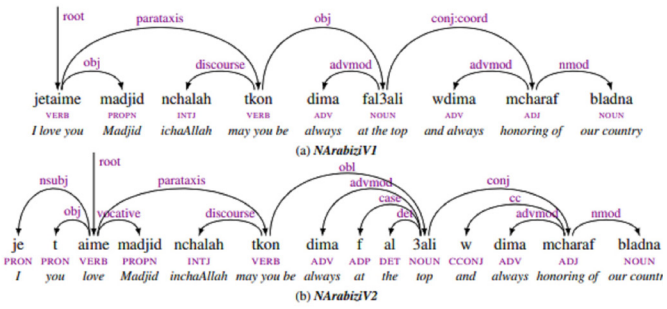


Figure 1: Illustration of an example from the NAarabizi treebank before and after the modifications.

Towards a Robust Detection of Language Model Generated Text: Is ChatGPT that Easy to Detect? | INRIA

Published on 12 July 2023

Author: Wissam Antoun, Virginie Mouilleron, Benoît Sagot, Djamé Seddah

Recent advances in natural language processing (NLP) have led to the development of large language models (LLMs) such as ChatGPT. This paper proposes a methodology for developing and evaluating ChatGPT detectors for French text, with a focus on investigating their robustness on out-of-domain data and against common attack schemes. The proposed method involves translating an English dataset into French and training a classifier on the translated data.

Data-Efficient French Language Modeling with CamemBERTa | INRIA

Published on 2 August 2023

Author: Wissam Antoun, Benoît Sagot, Djamé Seddah

In this paper, we introduce CamemBERTa, a French DeBERTa model that builds upon the DeBERTaV3 architecture and training objective. We evaluate our model's performance on a variety of French downstream tasks and datasets, including question answering, part-of-speech tagging, dependency parsing, named entity recognition, and the FLUE benchmark, and compare against CamemBERT, the state-of-the-art monolingual model for French.

Greedy routing optimisation in hyperbolic networks | ELTE

Published on 20 March 2023

Author: Bendegúz Sulyok, Gegely Palla

One of the motivations behind hyperbolic embedding is that optimal placement of the nodes in a hyperbolic space is widely thought to enable efficient navigation on top of the network. According to that, one of the measures that can be used to quantify the quality of different embeddings is given by the fraction of successful greedy paths following a simple navigation protocol based on the hyperbolic coordinates. In the present work, we develop an optimisation scheme for this score in the native disk representation of the hyperbolic space.

From Text to Source: Results in Detecting Large Language Model-Generated Content | INRIA

Published on 1 October 2023

Author: Wissam Antoun, Benoît Sagot, Djamé Seddah

This paper investigates "Cross-Model Detection," evaluating whether a classifier trained to distinguish between source LLM-generated and human-written text can also detect text from a target LLM without further training. The study comprehensively explores various LLM sizes and families, and assesses the impact of conversational fine-tuning techniques on classifier generalization. The research also delves into Model Attribution, encompassing source model identification, model family classification, and model size classification.

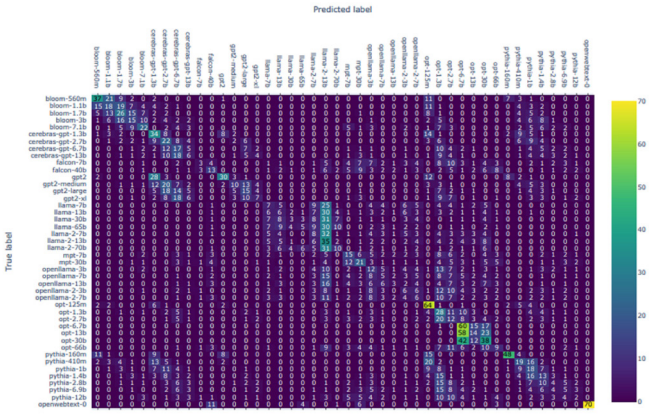


Figure 4: Normalized confusion matrix for Source Model Identification. 5-seed averaged and normalized by the predicted class support.

HURI: Hybrid user risk identification in social networks | CINI

Published on 29 July 2023

Author: Roberto Corizzo, Gianvito Pio, Emanuele Pio Barracchia, Antonio Pellicani, Nathalie Japkowicz, Michelangelo Ceci

In this paper, we propose HURI, a method for social network analysis that accurately classifies users as safe or risky, according to their behaviour in the social network. Specifically, the proposed hybrid approach leverages both the topology of the network of interactions and the semantics of the content shared by users, leading to an accurate classification also in the presence of noisy data, such as users who may appear to be risky due to the topic of their posts, but are actually safe according to their relationships.

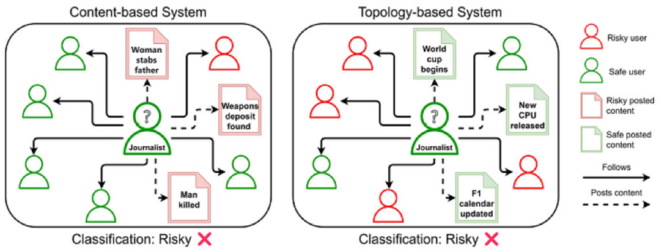


Fig. 1 A graphical representation of common misclassification errors made on a noisy user (e.g., a journalist). On the left, we show a misclassification error commonly made by content-based systems on users who post/interact with apparently risky content, even if he/she is linked with several safe users. On the right, we show a misclassification error commonly made by topology-based systems on users who establish more relationships with risky users than with safe users, even if he/she posts only safe content

Stay Tuned to Newsletter's Issue #8!

In issue #9, we will present the results from the project's second webinar, scheduled for January 2024. We will also provide information about the sixth CounteR technical meeting, paired with the fourth platform piloting session, due in February.

CounteR Consortium

ASSIST Software SRL, Romania

INSIKT Intelligence, Spain

Imagga Technologies LtD., Bulgaria

ICON Studios Ltd., Malta

Consorzio Interuniversitario Nazionale per l'Informatica, Italy

Institut national de Recherche en Informatique et Automatique, France

Eötvös Loránd University, Hungary

Università Cattolica del Sacro Cuore Academia, Italy

Malta Information Technology Law Association, Malta

European Institute Foundation, Bulgaria

Association les Militants du Savoir, France

ETICAS Research and Consulting, Spain

Elliniki Etairia Tilepoikinonion kai Tilematikon Efarmogon A.E.-Forthnet S.A., Greece

Ministério da Justiça, Portugal

Hochschule für den öffentlichen Dienst in Bayern, Germany

State Police Latvia, Latvia

Serviciul de Protectie si Paza, Romania

National Police General Directorate at the Ministry of Interior of the Republic of Bulgaria

Ministère de l'intérieur/ direction générale de la sécurité intérieure, France

YOUR COUNTER POINTS OF CONTACT ARE:

Mr. Catalin Trufin, Project Coordinator
ASSIST Software SRL
catalin.trufin@assist.ro

Mr. Borislav Mavrov, Dissemination manager
European Institute
bmavrov@europe.bg

https://counter-project.eu

https://twitter.com/CounteRH2020

https://www.linkedin.com/company/counter-project/

The CounteR project *has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no. 101021607.*