THE FIRST INTERNATIONAL CONFERENCE

ON

NATURAL LANGUAGE PROCESSING AND ARTIFICIAL INTELLIGENCE FOR CYBER SECURITY

NLPAICS'2024

P R O C E E D I N G S

Edited by:
Ruslan Mitkov, Saad Ezzini, Cengiz Acarturk, Tharindu Ranasinghe, Paul Rayson, Mo El-Haj, Ignatius Ezeani, Matthew Bradbury, Nouran Khallaf

Lancaster, United Kingdom
July 29 - 30, 2024
https://nlpaics.com/

THE FIRST INTERNATIONAL CONFERENCE
ON NATURAL LANGUAGE PROCESSING AND
ARTIFICIAL INTELLIGENCE FOR CYBER SECURITY

PROCEEDINGS

29–30 July, 2024
https://nlpaics.com/
ISBN: 978-1-86220-430-0

Message from the General Chair

In today's digital world, Cyber Security has emerged as a heightened priority for both individual users and organisations. As the volume of online information grows exponentially, traditional security approaches often struggle to identify and prevent evolving security threats. The inadequacy of conventional security frameworks highlights the need for innovative solutions that can effectively navigate the complex digital landscape for ensuring robust security. Natural Language Processing and Artificial Intelligence in Cyber Security have vast potential to significantly enhance threat detection and mitigation by fostering the development of advanced security systems for autonomous identification, assessment, and response to security threats in real-time. Recognising this challenge and the capabilities of Natural Language Processing (NLP) and Artificial Intelligence (AI) approaches to fortify Cyber Security (CS) systems, the First International Conference on Natural Language Processing and Artificial Intelligence for Cyber Security (NLPAICS'2024) serves as a gathering place for researchers in NLP and AI methods for Cyber Security. We invite contributions that present the latest NLP and AI solutions for mitigating risks in processing digital information.

This first-of-its-kind event covers a number of topics related to Cyber Security falling under (but not limited to) the following more general areas: Societal and Human Security and Safety; Speech Technology and Multimodal Investigations for Cyber Security; Data and Software Security; Human-Centric Security and Support; Anomaly Detection and Threat Intelligence; Systems and Infrastructure Security; Financial Cyber Security; Ethics, Bias, and Legislation in Cyber Security; Datasets and resources for Cyber Security Applications and Specialised Security Applications and Open Topics. It also features a Special Theme Track "Future of Cyber Security in the Era of LLMs and Generative AI".

We would like to thank all colleagues who made this unique international event possible. We would like to start by thanking all colleagues who submitted papers to NLPAICS'2024 and travelled to Lancaster to attend the event. We are grateful to all members of the Programme Committee for carefully evaluating all submissions (every submission was reviewed by 3 reviewers) and providing substantial feedback on all papers, helping the authors of accepted papers to improve and polish the final versions of their papers. A special thanks goes to all keynote speakers (Iva Gumnishka, Sevil Şen, Paolo Rosso and Jacques Klein). The role of the sponsors (Mind Bridge AI, Data Science Institute, Security Lancaster, UCREL) is acknowledged with gratitude.

Last but not least, we would like to use this paragraph to acknowledge the members of the Organising Committee whose dedication and efforts during the last 10 months made the organisation of this event possible. A big 'Thank you' goes to Prof Nigel Davies, Dr Saad Ezzini, Dr Tharindu Ranasinghe, Prof Paul Rayson, Dr Cengiz Acartürk, Dr Mo El-Haj, Dr Matthew Bradbury, Dr Ignatius Ezeani, Dr Amal Haddad Haddad, Dr Nouran Khallaf, Julia Carradus, Sofiia Denysiuk, and Isla Cambell.

Welcome to NLPAICS'2024 in Lancaster and we hope you will enjoy the event.

Prof Ruslan Mitkov
NLPAICS'2024 Conference Chair

Organisers and Sponsors

The First International Conference on Natural Language Processing and Artificial Intelligence for Cyber Security is organised by:

Conference Chair: Ruslan Mitkov, Lancaster University, UK

Programme Chairs:
- Cengiz Acartürk, Jagiellonian University, Poland
- Matthew Bradbury, Lancaster University, UK
- Mo El-Haj, Lancaster University, UK
- Paul Rayson, Lancaster University, UK

Sponsorship Chair:
- Saad Ezzini, Lancaster University, UK

Publicity Chair:
- Tharindu Ranasinghe, Lancaster University, UK

Publication Chair:
- Ignatius Ezeani, Lancaster University, UK

Social Programme Chair:
- Nouran Khallaf, Lancaster University, UK

Programme Committee:
- Hamza Alami, Sidi Mohamed Ben Abdellah University, Morocco
- Bharathi Raja Asoka Chakravarth, University of Galway, Ireland
- Enam Al-Wer, University of Essex, UK
- Pelin Angin, Middle East Technical University, Turkey
- Abdessamad Benlahbib, Sidi Mohamed Ben Abdellah University, Morocco
- Ismail Berrada, Mohammed VI Polytechnic University, Morocco
- Tegawende Bissyande, The University of Luxembourg, Luxembourg
- Alba Bonet Jover, University of Alicante, Spain
- Vaclav Brezina, Lancaster University, UK
- Georgina Brown, Lancaster University, UK
- Matthew Edwards, University of Bristol, UK
- Ahmed Elmesiry, London Metropolitan University, UK
- Sofia Ellina, Lancaster University, UK
- Ashraf Elnagar, University of Sharjah, United Arab Emirates
- Dan Fretwell, Lancaster University, UK
- Claire Hardaker, Lancaster University, UK
- Amit Kumar Jaiswal, University of Surrey, UK
- Daisy Monika Lal, Lancaster University, UK
- Isabel Espinosa Zaragoza, University of Alicante, Spain
- Hassane Essafi, CEA, France
- Pinar Gurkan, Researcher, Turkey

- Hongmei He, University of Salford, UK
- Hansi Hettiarachchi, Birmingham City University, UK
- Nouran Khallaf, Lancaster University, UK
- Jacques Klein, University of Luxembourg, Luxembourg
- Henrik Legind Larsen, Legind Technologies AS, Denmark
- Wanpeng Li, University of Aberdeen, UK
- Juan Carlos Nieves Sanchez, Umeå University, Sweden
- Eugenio Martínez Cámara, University of Jaen, Spain
- Scott Piao, Lancaster University, UK
- Lena Podoletz, Lancaster University, UK
- Pattabhi RK Rao, Anna University Chennai, India
- Nasredine Semmar, CEA/University of Paris-Saclay, France
- Rui Sousa Silva, University of Porto, Portugal
- Mark Stamp, San Jose State University, USA
- Sandra Kübler, Indiana University, USA
- Simon Parkinson, Huddersfield University, UK
- Tharindu Ranasinghe, Lancaster University, UK
- Ricardo J. Rodríguez, University of Zaragoza, Spain
- Sriram Sagi, NetApp, California, USA
- Sevil Şen, Hacettepe University, Turkey
- Cihan Varol, Sam Houston State University, USA
- Alfonso Ureña, University of Jaen, Spain
- Sule Yildirim Yayilgan, Norwegian University of Science and Technology, Norway
- Wajdi Zaghouani, Hamad Bin Khalifa University, Qatar
- Marcos Zampieri, George Mason University, USA
- Ayah Zirikly, John Hopkins University, USA

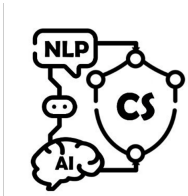
Additional Reviewers:
- Shrikant Malviya, Durham University, UK
- Atul Kumar Ojha, University of Galway, Ireland
- Adarsh Prasad Behera, IMDEA Networks Institute, Spain
- Vladimir Petrov, Lancaster University, UK
- Markus Barrot, Lancaster University, UK

Advisory Committee:
- Cengiz Acartürk, Jagiellonian University, Poland
- Simon Cook, Lancaster University, UK
- Nigel Davies, Lancaster University, UK
- Basil Germond, Lancaster University, UK
- Amal Haddad Haddad, University of Granada, Spain
- Raquel Lázaro Gutiérrez, Universidad de Alcalá, Spain
- Rafael Muñoz Guillena, University of Alicante, Spain
- Sandra Kübler, Indiana University, USA
- Tony McEnery, Lancaster University, UK
- Manuel Palomar, University of Alicante, Spain
- Daniel Prince, Lancaster University, UK
- Nicholas Race, Lancaster University, UK
- Paul Rayson, Lancaster University, UK

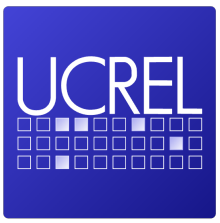
- Tharindu Ranasinghe, Lancaster University, UK
- Neraj Suri, Lancaster University, UK
- Alfonso Ureña-López, University of Jaen, Spain
- Marcos Zampieri, George Mason University, USA

Organising Committee:

- Abeer Alotaibi, Lancaster University, UK
- Maram Alharbi, Lancaster University, UK
- Hilary Barraclough, Lancaster University, UK
- Andreea Beca, Lancaster University, UK
- Julia Carradus, Lancaster University, UK
- Sofiia Denysiuk, Lancaster University, UK
- Sofia Ellina, Lancaster University, UK
- Wenjing Jiang, Lancaster University, UK
- Nouran Khallaf, Lancaster University, UK
- Daisy Monika Lal, Lancaster University, UK
- Henry Oldroyd, Lancaster University, UK
- Lena Podoletz, Lancaster University, UK
- Vladimir Petrov, Lancaster University, UK
- Yafei Zhu, Lancaster University, UK



NLPAICS'2024 Sponsors



Security Lancaster | Lancaster University

Data Science Institute | Lancaster University

School of Computing & Communications | Lancaster University

Table of Contents

Monday, July 29, 2024

8:30–9:10 *Registration*

9:10–9:20 *Opening*
Nigel Davies, Head of the School of Computing and Communications

9:20–10:10 *Keynote speech 1* (40 min presentation; 10 min Q & A; Introduction: Ruslan Mitkov)
Paolo Rosso – Beyond fake news in disinformation detection: analysis of narratives of conspiracy theories

10:10–11:00 Session 1: Ethics and bias (Session Chair: Eugenio Martínez Camara)

10:10–10:35 *Predatory Publication of AI-Generated Research Papers*
Lizzie Burgiss, Ben Tatum, Christopher Henshaw, Madison Boswell, and Alan Michaels

10:35–11:00 *Explainability of Machine Learning Approaches in Forensic Linguistics: A Case Study in Geolinguistic Authorship Profiling*
Dana Roemling, Yves Scherrer, and Aleksandra Miletić

11:00–11:20 *Morning coffee break*

11:20–12:35 **Session 2: Software and vulnerabilities** (Session Chair: Matthew Bradbury)

11:20–11:45 *Metric-Oriented Pretraining of Neural Source Code Summarisation Transformers to Enable more Secure Software Development*
Jesse Phillips, Mo El-Haj, and Tracy Hall

11:45–12:10 *Comprehensive threat analysis and systematic mapping of CVEs to MITRE framework*
Stefano Simonetto and Peter Bosch

12:10–12:35 *Predicting Software Vulnerability Trends with Multi-Recurrent Neural Networks: A Time Series Forecasting Approach*
Abanisenioluwa K. Orojo, Webster C. Elumelu, and Oluwatamilore O. Orojo

12:35–1:00 *Measuring the Effect of Induced Persona on Agenda Creation in Language-based Agents for Cyber Deception*
Lewis Newsham, Daniel Prince, and Ryan Hyland

1:00–2:00 *Lunch break*

2:00–2:50 *Keynote speech 2* (Introduction: Hansi Hettiarachchi)
Sevil Sen – AI versus AI: The Relentless Cyber Security Arms Race

2:50–4:05 **Session 3: Spam and phishing** (Session Chair: Lena Podoletz)

2:50–3:15 *Comparative Analysis of Natural Language Processing Models for Malware Spam Email Identification*
Francisco Jáñez-Martino, Eduardo Fidalgo, Rocío Alaiz-Rodríguez, Andrés Carofilis, and Alicia Martínez-Mendoza

3:15–3:40 *SpamClus: An Agglomerative Clustering Algorithm for Spam Email Campaigns Detection*
Daniel Díaz, Wesam Al-Nabki, Laura Fernández-Robles, Enrique Alegre, Eduardo Fidalgo, and Alicia Martínez-Mendoza

3:40–4:05 *LSTM-PSO: NLP-based model for detecting Phishing Attacks*
Abdulrahman A. Alshdadi

4:05–4:25 *Afternoon coffee break*

4:25–4:55 *Sponsor (Mind Bridge AI) presentation* (Introduction: Saad Ezzini)
CodeAgent - Collaborative Agents for Software Engineering
Daniel Tang

4:55–6:35 **Session 4: Fake news, privacy and NLP challenges** (Session Chair: Rafael Muñoz Guillena)

4:55–5:20 *The Influence of the Perplexity Score in the Detection of Machine-generated Texts*
Alberto José Gutiérrez Megías, L. Alfonso Ureña-López, and Eugenio Martínez Cámara

5:20–5:45 *Variation between Credible and Non-Credible News Across Topics*
Emilie Francis

5:45–6:10 *Can LLMs assist with Ambiguity? A Quantitative Evaluation of various Large Language Models on Word Sense Disambiguation*
Deshan Koshala Sumanathilaka, Nicholas Micallef, and Julian Hough

6:10–6:35 *Privacy Preservation in Federated Market Basket Analysis using Homomorphic Encryption*
Sameeka Saini and Durga Toshniwal

19:45–22:30 *Conference dinner*

Tuesday, July 30, 2024

9:00–9:20 *Coffee*

9:20–10:10 *Keynote speech 3* (Introduction: Tharindu Ranasinghe)
Iva Gumnishka - Red Teaming: Trustworthy AI through diverse human testing

10:10–11:00 Session 5: Anomaly Detection and Threat Intelligence (Session Chair: Ignatius Ezeani)

10:10–10:35 *WAVE-27K: Bringing together CTI Sources for Enhanced Threat Intelligence Models*
Felipe Castaño, Amaia Gil-Lerchundi, Raul Orduna-Urrutia, Eduardo Fidalgo Fernandez, and Rocío Alaiz-Rodríguez

10:35–11:00 *Human-in-the-loop Anomaly Detection and Contextual Intelligence for Enhancing Cybersecurity Management*
Thomas Schaberreiter, Jerry Andriessen, Cinzia Cappiello, Alex Papanikolaou, and Mirjam Pardijs

11:00–11:20 *Morning coffee break*

11:20–1:00 **Session 6: Hate speech and harmful content** (Session Chair: Claire Hardaker)

11:20–11:45 *Is it Offensive or Abusive? An Empirical Study of Hateful Language Detection of Arabic Social Media Texts*
Salim Al Mandhari, Mo El-Haj, and Paul Rayson

11:45–12:10 *The Elsagate Corpus: Characterising Commentary on Alarming Video Content*
Panagiotis Soustas and Matthew Edwards

12:10–12:35 *Abusive Speech Detection in Serbian using Machine Learning*
Danka Jokić, Ranka Stanković, and Branislava Šandrih Todorović

12:35–1:00 *Fighting Cyber-malice: A Forensic Linguistics Approach to Detecting AI-generated Malicious Texts*
Rui Sousa-Silva

1:00–2:00 *Lunch break*

2:00–2:50 *Keynote speech 4: Jacques Klein* (Introduction: Saad Ezzini)

2:50–4:05 **Session 7: Threats and vulnerabilities** (Session Chair: Ashley Fraser)

2:50–3:15 *Deciphering Cyber Threats: A Unifying Framework with GPT-3.5, BERTopic and Feature Importance*
Chun Man Tsang, Tom Bell, Antonios Gouglidis, and Mo El-Haj

3:15–3:40 *CECILIA: Enhancing CSIRT Effectiveness with Transformer-Based Cyber Incident Classification*
Juan Jose Delgado Sotes, Alicia Martinez Mendoza, Andres Carofilis Vasco, Eduardo Fidalgo Fernandez, and Enrique Alegre Gutierrez

3:40–4:05 *U-BERTopic: An Urgency-Aware BERT-Topic Modeling Approach for Detecting CyberSecurity Issues via Social Media*
Majed Albarrak, Gabriele Pergola, and Arshad Jhumka

4:05–4:25 *Afternoon coffee break*

4:25–6:05 **Session 8: LLM and vulnerabilities** (Session Chair: Mo El-Haj)

4:25–4:50 *A Proposal Framework Security Assessment for Large Language Models*
Daniel Mendonça Colares, Raimir Holanda Filho, and Luis Borges Gouveia

4:50–5:15 *Not Everything Is Online Grooming: False Risk Finding in Large Language Model Assessments of Human Conversations*
Ellie Prosser and Matthew Edwards

5:15–5:40 *Redacted Contextual Question Answering with Generative Large Language Models*
Jacob Lichtefeld, Joe A. Cecil, Alex Hedges, Jeremy Abramson, and Marjorie Freedmann

5:40–6:05 *Unlocking LLMs Capabilities: Addressing Scarce Data and Inherent Bias Challenges in Mental Health and Therapeutic Counselling*
Vivek Kumar, Pushpraj Singh Rajwat, Giacomo Medda, Eirini Ntoutsi, and Diego Reforgiato Recupero

6:05–6:15 *Closing*

19:30–22:00 *Networking event*