THE SECOND INTERNATIONAL CONFERENCE

ON

NATURAL LANGUAGE PROCESSING AND ARTIFICIAL INTELLIGENCE FOR CYBER SECURITY

NLPAICS'2026

P R O C E E D I N G S

Edited by:
Ruslan Mitkov, Rafael Muñoz, Elena Lloret, Tharindu Ranasinghe,
Ernesto L. Estevanell-Valladares, Salima Lamsiyah, Andrés Montoyo,
Saad Ezzini

Alicante, Spain
June 11–12, 2026

https://nlpaics2026.gplsi.es/

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

PROCEEDINGS

11–12 June, 2026
Alicante, Spain
https://nlpaics2026.gplsi.es/

NLPAICS'2026 is organised by the GPLSI research group (Grupo de Procesamiento del Lenguaje y Sistemas de Información) of the University of Alicante, with the support of the University Institute for Computing Research (Instituto Universitario de Investigación Informática, IUII).

The conference has been organised within the framework of the CIDEGENT project (ref. CIDEXG/2023/13) of the Generalitat Valenciana. The organisers gratefully acknowledge the support of the Generalitat Valenciana, through the Conselleria de Educación, Cultura, Universidades y Empleo, and of the University of Alicante.

ISBN: 978-84-1302-372-4

Message from the Conference Chairs

Welcome to Alicante, and to the Second International Conference on Natural Language Processing and Artificial Intelligence for Cyber Security (NLPAICS'2026).

Cyber Security matters more each year. As more of our lives moves online, many of the threats we face now take the form of language: phishing emails, fake news, machine-generated text, and synthetic voices used to deceive. Natural Language Processing and Artificial Intelligence have become some of our best tools for detecting and countering these threats, and the papers in this volume show how much the two communities can achieve when they work together.

NLPAICS began in Lancaster in 2024, bringing the language and security communities together for the first time. It was warmly received, and we are very pleased that the second edition now takes place at the University of Alicante.

The papers cover a broad range of topics, from disinformation, online abuse and machine-generated text to digital forensics, malware analysis, threat intelligence, and the ethics and regulation of AI in security. As in Lancaster, the conference includes a special theme track on the future of cyber security in the era of large language models and generative AI. This year we are also pleased to host a student workshop and to present awards for the best papers.

A conference is the work of many people. We thank the authors who submitted their work and came to Alicante to present it, and the members of the Programme Committee, whose reviews helped improve every paper. We are grateful to the Generalitat Valenciana and the University of Alicante for their support, and to the GPLSI research group and the Instituto Universitario de Investigación Informática for hosting the event. Above all, we thank the Organising Committee for their hard work over many months.

We hope you enjoy the conference and find some time for Alicante while you are here. Welcome to NLPAICS'2026.

Ruslan Mitkov and Rafael Muñoz
NLPAICS'2026 Conference Chairs

Message from the Programme Chairs

The proceedings before you bring together the papers accepted for the second edition of NLPAICS, the result of several months of work by their authors and by the Programme Committee.

Every submission was reviewed by at least three members of the Programme Committee, who gave their time without reward and often under time pressure. We are grateful to them for the care they took, and to the authors, who responded to the reviews thoughtfully and improved their papers as a result.

The conference again includes the special theme track introduced in Lancaster, "Future of Cyber Security in the Era of LLMs and Generative AI". With large language models now in everyday use, the questions it asks about safety, bias, and the responsible use of these systems have become harder to avoid, and we were pleased by how readily the community engaged with them.

We also thank our keynote speakers, Preslav Nakov, Tharindu Ranasinghe, Salima Lamsiyah and Saad Ezzini, for accepting our invitation.

These proceedings will be archived in the ACL Anthology, and we hope they prove useful to the community working on NLP and AI for Cyber Security.

Tharindu Ranasinghe and Elena Lloret
NLPAICS'2026 Programme Chairs

Organising Committee

Conference Chairs
Ruslan Mitkov, University of Alicante, Spain
Rafael Muñoz, University of Alicante, Spain

Programme Chairs
Tharindu Ranasinghe, Lancaster University, UK
Elena Lloret, University of Alicante, Spain

Publication Chair
Ernesto L. Estevanell-Valladares, University of Alicante, Spain

Sponsorship Chair
Andrés Montoyo, University of Alicante, Spain

Student Workshop Chair and Best Paper Chair
Salima Lamsiyah, University of Luxembourg, Luxembourg

Best Student Paper Award Chair
Saad Ezzini, King Fahd University of Petroleum and Minerals, Saudi Arabia

Publicity Chair
Beatriz Botella, University of Alicante, Spain

Social Programme Chair
Alba Bonet, University of Alicante, Spain

Local Organisation and Support Team
Sergio Castillo Blanco, Santander Industrial University, Spain
Iván Martínez Murillo, University of Alicante, Spain
Raúl García Cerdá, University of Alicante, Spain
Silvia Gargova, University of Alicante, Spain
Nouran Khallaf, University of Leeds, UK

Webmaster
Sergio Espinosa Zaragoza, University of Alicante, Spain

Advisory Committee
Manuel Palomar, CENID and GPLSI, University of Alicante, Spain
Rafael Muñoz, University of Alicante, Spain
Andrés Montoyo, University of Alicante, Spain

Programme Committee

Cengiz Acartürk, Jagiellonian University, Poland
Sedat Akleylek, University of Tartu, Estonia
Amneh Al Abdi, Ajloun National University, Jordan
Hind Alatawi, University of Tabuk, Saudi Arabia
Angela Almela Sánchez-Lafuente, University of Murcia, Spain

Houdaifa Atou, Mohammed VI Polytechnic University, Morocco
Abdessamad Benlahbib, Sidi Mohamed Ben Abdellah University, Morocco
Ismail Berrada, Sidi Mohamed Ben Abdellah University, Morocco
Matthew Bradbury, Lancaster University, UK
Salmane Chafik, Mohammed VI Polytechnic University, Morocco
Matthew Edwards, University of Bristol, UK
Ashraf Elnagar, University of Sharjah, United Arab Emirates
Marie Escribe, University of Valencia, Spain
Eduardo Fidalgo, University of León, Spain
Dan Fretwell, Lancaster University, UK
Basil Germond, Lancaster University, UK
Amal Haddad Haddad, University of Granada, Spain
Ogtay Hasanov, Baku Engineering University, Azerbaijan
Hongmei He, University of Salford, UK
Hansi Hettiarachchi, Lancaster University, UK
Ozkan Kilic, Cisco Systems Inc., USA
Jacques Klein, University of Luxembourg, Luxembourg
Gražina Korvel, Vilnius University, Lithuania
Sandra Kübler, Indiana University Bloomington, USA
Vivek Kumar, University of the Bundeswehr München, Germany
Daisy Lal, Lancaster University, UK
Henrik Larsen, Aalborg University, Denmark
Francisco Javier Lima Florido, University of Málaga, Spain
Shrikant Malviya, Durham University, UK
Eugenio Martínez Cámara, University of Jaén, Spain
Alicia Martínez-Mendoza, University of León, Spain
Iván Martínez Murillo, University of Alicante, Spain
Ilker Ozcelik, Eskisehir Osmangazi University, Turkey
Lena Podoletz, Lancaster University, UK
Daniel Prince, Lancaster University, UK
Nasredine Semmar, CEA / University of Paris-Saclay, France
Sevil Sen, Hacettepe University, Turkey
Rui Sousa-Silva, University of Porto, Portugal
Alina Trubnikova, University of Granada, Spain
Ozgur Ural, Embry-Riddle Aeronautical University, USA
Alfonso Ureña López, University of Jaén, Spain
Rafael Valencia García, University of Murcia, Spain
Wajdi Zaghouani, Northwestern University Qatar
Marcos Zampieri, George Mason University, USA
Xiaojing Zhao, Hong Kong Polytechnic University, Hong Kong

Table of Contents

Exploring Cross-Lingual Transfer in Transformer-Based Fraud Detection Models
Iván Martínez Murillo, Robiert Sepúlveda-Torres and Juan Pablo Consuegra-Ayala 1

Grieve-JA: A Psycholinguistic Dictionary for Grievance-Fuelled Language in Japanese
Emilie Francis 11

Pragmatic Profiling for Disinformation Detection: An Exploratory Analysis of Stylistic Features in Spanish News
Alba Pérez-Montero, María Miró Maestre, Elena Lloret and Paloma Moreda 25

Transformer-Assisted LLM-Based Source Code Summarisation: to Enable More Secure Software Development
Jesse Phillips, Tracy Hall, Paul Rayson and Mo El-Haj 36

Code Without Context: Can We Trust LLMs to Test Software from Informal Descriptions?
Amneh Al Abdi and Saad Ezzini 46

Fine-Tuning Small Language Models for Cybersecurity: Data Ordering, Knowledge Distillation, and the Educator Effect
Ozkan Kilic, Raja Soundaramourty and Ramu Chenchaiah 55

Large-Scale Multilingual SMS Fraud Detection For Telecom Networks
Orlando Amaral Cejas, Yuejun Guo and Qiang Tang 64

Command-Line Obfuscation Detection in Real-World Telemetry under Extreme Class Imbalance
Vojtěch Outrata, Barbora Štěpánková, Michael Adam Polák and Martin Kopp 74

OBSIDIAN: An OSINT-Driven NLP Framework for Detecting Cyber-Physical Threats in Arabic Social Media
Abdullah Saeed Almalki, Salmane Chafik and Saad Ezzini 88

Bloc-Conditional Event States: Measuring Cross-Coverage Divergence for Threat-Intelligence Analysis
Maryam Fooladi and Federico Bottino 98

CSULoRA: Closest Safe Update Low-Rank Adaptation
Oleksandr Marchenko Breneur, Adelaide Danilov, Aria Nourbakhsh and Salima Lamsiyah 103

LLM-based Defense Against Adversarial Abstracts in ML/AI Conference Reviewer Assignments
Mohamed Omar Cherif, Christophe Cerisara and Julien Falgas 113

From Detection to Attribution: Forensic Linguistics and Adversarial Red Teaming as Complementary Responses to LLM Misuse
Rui Sousa-Silva 122

The Human Attack Surface: Detecting Psychological Vulnerabilities to Cyber Threats using AI and Social Media
Aadil Gani Ganie and Saad Ezzini 134

A Neuro-Cyber Exploitation and Reconnaissance Taxonomy (NeuroCERT) for Human-Centric Cybersecurity
Cengiz Acarturk, Melike Caglayan, Ece Caglayan and Anna Maria Wilkosz 143

Analysing Self-Harm Representations in Language Models: a Cross-Architecture Study
Luis Espinosa Anke and Carla Perez Almendros 156

A Linguistic Analysis of Prompt Injection in Large Language Models
Priscilla Adenuga 163

How Well Do Commodity Text-to-Speech Systems Evade Acoustic Perturbation Detection? A Multi-Engine Evaluation Across 21 Languages
Anatoly Marchenko 171

Semantic Clustering of Obfuscated Command-Line Detections for Alert Reduction
Barbora Štěpánková, Vojtěch Outrata and Martin Kopp 176

Judging the LLM Judges: A Human-Centric Validation of LLM-Generated Training Data for Software Retrieval
Ogtay Hasanov and Saad Ezzini 184

Exploring the Manifestation of Schwartz's Basic Human Values in Large Language Models
Ryan Hyland, Lewis Newsham and Daniel Prince 189

From Decision Tree to Detection Pipeline: Formalizing van Dijk's Socio-Cognitive Framework for Automated Anti-Language Identification in RICO Transcripts
Elena Morandini 204

When Privacy Helps: Pseudonymisation as a Strategy for Improved Cyber Incident Classification
Loya C. Haughton, Eduardo Fidalgo, Rocío Alaiz-Rodríguez, Manuel Castejón-Limas and Laura Fernández-Robles 215

Does Hate Transfer? Cross-Lingual Generalisation of Offensive Content Detection Across Indic Languages
Purandhar M. Reddy and Sara Renjit 228

LLMs in the Enterprise: A Systematic Review of Security Architectures for RAG-Augmented Chatbots
Fatimah Alali, Haya Aldawsari, Saad Ezzini and Sajjad Mahmood 232

Partner Attribution Bias in LLM-Assisted Export Control Screening
Salem Alotaibi, Alexei Lisitsa and Antony McCabe 244

Privacy VITA: a new multilingual and multimodal annotated video corpus to evaluate anonymization systems
Jorge Rico, Sofia Contreras, Enrique Manjavacas Arevalo, Maria Viana, Ruben Pérez-Ramón, Maria Luisa Izquierdo, María José Vilella, Jaime Corton, Silvia Rodriguez, Fernando Espinza, Rafael Ginard, César Pérez, Jesús Arias, Richard Cook, Pablo Regodón, Manuel Moyano, Ricardo Heredia, Lara De Santos, Pierre Plaza and Jacqueline González 253

Conference Program

Thursday, June 11, 2026

10:15–11:15 **Session S1: Session 1: Multilingual & Cross-Lingual NLP in Security Applications**
Chair: Ernesto Estevanell

10:15–10:35 *Exploring Cross-Lingual Transfer in Transformer-Based Fraud Detection Models*
Iván Martínez Murillo, Robiert Sepúlveda-Torres and Juan Pablo Consuegra-Ayala

10:35–10:55 *Grieve-JA: A Psycholinguistic Dictionary for Grievance-Fuelled Language in Japanese*
Emilie Francis

10:55–11:15 *Pragmatic Profiling for Disinformation Detection: An Exploratory Analysis of Stylistic Features in Spanish News*
Alba Pérez-Montero, María Miró Maestre, Elena Lloret and Paloma Moreda

12:45–13:45 **Session S2: Session 2: Securing the Software Development Lifecycle**
Chair: Hansi Hettiarachchi

12:45–13:05 *Transformer-Assisted LLM-Based Source Code Summarisation: to Enable More Secure Software Development*
Jesse Phillips, Tracy Hall, Paul Rayson and Mo El-Haj

13:05–13:25 *Code Without Context: Can We Trust LLMs to Test Software from Informal Descriptions?*
Amneh Al Abdi and Saad Ezzini

13:25–13:45 *Fine-Tuning Small Language Models for Cybersecurity: Data Ordering, Knowledge Distillation, and the Educator Effect*
Ozkan Kilic, Raja Soundaramourty and Ramu Chenchaiah

Thursday, June 11, 2026 (continued)

14:45–16:00 **Session S3: Session 3: Threat Detection in the Wild: Messages, Telemetry & Social Media**
Chair: Saad Ezzini

14:45–15:05 *Large-Scale Multilingual SMS Fraud Detection For Telecom Networks*
Orlando Amaral Cejas, Yuejun Guo and Qiang Tang

15:05–15:25 *Command-Line Obfuscation Detection in Real-World Telemetry under Extreme Class Imbalance*
Vojtěch Outrata, Barbora Štěpánková, Michael Adam Polák and Martin Kopp

15:25–15:45 *OBSIDIAN: An OSINT-Driven NLP Framework for Detecting Cyber-Physical Threats in Arabic Social Media*
Abdullah Saeed Almalki, Salmane Chafik and Saad Ezzini

15:45–16:00 *Bloc-Conditional Event States: Measuring Cross-Coverage Divergence for Threat-Intelligence Analysis*
Maryam Fooladi and Federico Bottino

16:30–17:30 **Session S4: Session 4: LLM Safety & Adversarial Robustness**
Chair: Houdaifa Atou

16:30–16:50 *CSULoRA: Closest Safe Update Low-Rank Adaptation*
Oleksandr Marchenko Breneur, Adelaide Danilov, Aria Nourbakhsh and Salima Lamsiyah

16:50–17:10 *LLM-based Defense Against Adversarial Abstracts in ML/AI Conference Reviewer Assignments*
Mohamed Omar Cherif, Christophe Cerisara and Julien Falgas

17:10–17:30 *From Detection to Attribution: Forensic Linguistics and Adversarial Red Teaming as Complementary Responses to LLM Misuse*
Rui Sousa-Silva

Friday, June 12, 2026

10:00–11:00 **Session S5: Session 5: Human-Centric Cybersecurity & Wellbeing**
Chair: Iván Martínez-Murillo

10:00–10:20 *The Human Attack Surface: Detecting Psychological Vulnerabilities to Cyber Threats using AI and Social Media*
Aadil Gani Ganie and Saad Ezzini

10:20–10:40 *A Neuro-Cyber Exploitation and Reconnaissance Taxonomy (NeuroCERT) for Human-Centric Cybersecurity*
Cengiz Acarturk, Melike Caglayan, Ece Caglayan and Anna Maria Wilkosz

10:40–10:55 *Analysing Self-Harm Representations in Language Models: a Cross-Architecture Study*
Luis Espinosa Anke and Carla Perez Almendros

11:30–12:30 **Session S6: Session 6: Detecting Attacks & Manipulation**
Chair: Alicia Picazo

11:30–11:45 *A Linguistic Analysis of Prompt Injection in Large Language Models*
Priscilla Adenuga

11:45–12:00 *How Well Do Commodity Text-to-Speech Systems Evade Acoustic Perturbation Detection? A Multi-Engine Evaluation Across 21 Languages*
Anatoly Marchenko

12:00–12:15 *Semantic Clustering of Obfuscated Command-Line Detections for Alert Reduction*
Barbora Štěpánková, Vojtěch Outrata and Martin Kopp

12:15–12:30 *Judging the LLM Judges: A Human-Centric Validation of LLM-Generated Training Data for Software Retrieval*
Ogtay Hasanov and Saad Ezzini

Friday, June 12, 2026 (continued)

12:30–14:00 **Session STU: Student Research Session**
Chair: Salima Lamsiyah

12:30–12:50 *Exploring the Manifestation of Schwartz's Basic Human Values in Large Language Models*
Ryan Hyland, Lewis Newsham and Daniel Prince

12:50–13:10 *From Decision Tree to Detection Pipeline: Formalizing van Dijk's Socio-Cognitive Framework for Automated Anti-Language Identification in RICO Transcripts*
Elena Morandini

13:10–13:30 *When Privacy Helps: Pseudonymisation as a Strategy for Improved Cyber Incident Classification*
Loya C. Haughton, Eduardo Fidalgo, Rocío Alaiz-Rodríguez, Manuel Castejón-Limas and Laura Fernández-Robles

13:30–13:45 *Does Hate Transfer? Cross-Lingual Generalisation of Offensive Content Detection Across Indic Languages*
Purandhar M. Reddy and Sara Renjit

16:30–17:30 **Session S7: Session 7: Enterprise & Regulated AI: Security, Compliance & Privacy**
Chair: Rui Sousa-Silva

16:30–16:50 *LLMs in the Enterprise: A Systematic Review of Security Architectures for RAG-Augmented Chatbots*
Fatimah Alali, Haya Aldawsari, Saad Ezzini and Sajjad Mahmood

16:50–17:10 *Partner Attribution Bias in LLM-Assisted Export Control Screening*
Salem Alotaibi, Alexei Lisitsa and Antony McCabe

17:10–17:25 *Privacy VITA: a new multilingual and multimodal annotated video corpus to evaluate anonymization systems*
Jorge Rico, Sofia Contreras, Enrique Manjavacas Arevalo, Maria Viana, Ruben Pérez-Ramón, Maria Luisa Izquierdo, María José Vilella, Jaime Corton, Silvia Rodriguez, Fernando Espinza, Rafael Ginard, César Pérez, Jesús Arias, Richard Cook, Pablo Regodón, Manuel Moyano, Ricardo Heredia, Lara De Santos, Pierre Plaza and Jacqueline González