LREC 2026

The Sixth Workshop on Resources and ProcessIng of linguistic, para-linguistic and extra-linguistic Data from people with various forms of cognitive/psychiatric/developmental impairments (RaPID@MENTAL.ai) @ LREC 2026

Workshop Proceedings

Editors
Dimitrios Kokkinakis, Charalambos Themistocleous, Gaël Dias, Kathleen C. Fraser, Sebastião Pais, Fredrik Öhman

12 May 2026

Proceedings of Resources and ProcessIng of linguistic, para-linguistic and extra-linguistic Data from people with various forms of cognitive/psychiatric/developmental impairments - In collaboration with the MENTAL.ai project (RaPID-6@MENTAL.ai 2026) @ LREC 2026

ISBN 978-2-493814-59-3

Message from the General Chair

Welcome to the LREC 2026 Workshop on *"Resources and ProcessIng of linguistic, para-linguistic and extra-linguistic Data from people with various forms of cognitive/psychiatric/developmental impairments"*, this year in a cooperation with the *"MENTAL.ai consortium"* (RaPID-6@MENTAL.ai). This volume documents the Proceedings of the RaPID-6@MENTAL.ai Workshop, held on Tuesday afternoon, May 12th, 2026, as part of the 15th edition of the LREC 2026 conference *"the International Conference on Language Resources and Evaluation"*.

RaPID-6@MENTAL.ai aims to be an interdisciplinary forum for researchers to share information, findings, methods, models and experience on the collection and processing of data produced by people with various forms of mental, cognitive, neuropsychiatric, or neurodegenerative impairments, such as aphasia, dementia, autism, bipolar disorder, Parkinson's disease, or schizophrenia. Like the previous five editions, the RaPID-6@MENTAL.ai workshop's focus is on creation, processing, and application of data resources from individuals at various stages of these impairments and with varying degrees of severity. Creation of resources includes e.g. annotation, description, analysis, and interpretation of linguistic, paralinguistic and extra-linguistic data (such as spontaneous spoken language, transcripts, eye tracking measurements, wearable and sensor data, etc). Processing is done to identify, extract, correlate, evaluate and disseminate various linguistic or multimodal phenotypes and measurements, which then can be applied to aid diagnosis, monitor the progression, or predict individuals at risk.

RaPID-6@MENTAL.ai invited submissions of papers in all of the aforementioned research areas, particularly emphasizing the multidisciplinary aspects of processing such data and the interplay between clinical, nursing, medical sciences, language technology, computational linguistics, natural language processing/artificial intelligence (NLP/AI), and computer science. The workshop serves as a catalyst for discussing several ongoing research questions that drive both current and future research endeavours, by bringing together researchers from diverse communities. The workshop invited papers describing original research, preferably presenting substantial and completed work, while also welcoming contributions such as negative results, interesting application nuggets, software packages / tools / platforms, small works, or works in progress. It stimulated discussions on various ongoing research questions and challenges by uniting researchers from different communities. We extend our gratitude to the members of the Scientific Program Committee (SPC) for their diligent efforts in reviewing and evaluating all submissions. Each submission received between 2 to 4 reviews, aiding authors in revising and improving their papers accordingly.

There were 12 contributions accepted for the workshop.

Keynote speakers of RaPID-6@MENTAL.ai were:

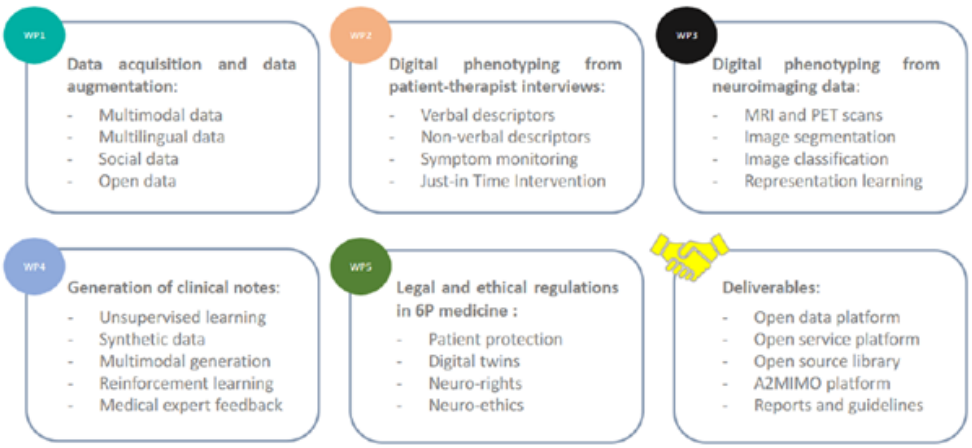
- **Brian MacWhinney**, Teresa Heinz Professor of Cognitive Psychology, Carnegie Mellon University, USA: *TalkBank resources for studying functional communication in language disorders - a survey*; and,
- **Sunny X. Tang**, Associated professor of psychiatry, MD, Feinstein Institutes for Medical Research, Northwell Health, USA: *Speech and language markers for clinical applications in psychiatric disorders*.

Workshop URL: https://spraakbanken.gu.se/en/rapid-2026.

The MENTAL.ai - Summary and Objectives

The MENTAL.AI (actually the MENTAL.AI@CaeSAR*) project aims to develop new AI–based approaches to improve the detection, monitoring, diagnosis, and management of psychiatric disorders and neurodegenerative diseases, such as depression, schizophrenia, and Alzheimer's disease. These conditions currently represent a major public health challenge, affecting millions of people worldwide and having a significant impact on patients, their families, and healthcare systems. The project's goal is to contribute to the development of the 6P medicine (personalized, predictive, preventive, participatory, proof-based, and pathway–oriented) by leveraging the advanced analytical capabilities of AI. By combining various data sources, including text, speech, vision, neuroimaging, behavioral and environmental signals, the project seeks to identify digital markers (digital phenotyping) that allow for earlier symptom detection and a better understanding of disease progression.
The project is based on several complementary research axes. The first involves creating a multimodal and multilingual dataset from clinical interviews conducted via a dedicated digital platform A2MIMO**, allowing for the collection of textual, speech, and behavioral and environmental information (Working Package, WP1). The second aims to develop methods for automatically analyzing patient-therapist interactions to identify linguistic, emotional, or cognitive markers characteristic of specific disorders (WP2). A third axis focuses on analyzing neuroimaging data (MRI, PET) combined with structured tabular information using advanced machine learning techniques to improve diagnostic accuracy, predict disease progression and evaluate patient risks (WP3). The project also explores the use of generative language models to automatically produce clinical reports from medical interviews or imaging data, thereby reducing the administrative burden on healthcare professionals (WP4). Finally, special attention is given to the ethical and legal issues related to the use of AI in the sensitive field of mental health, particularly data protection, patient consent and autonomy, the respect of patients' psychological integrity, liability and accountability, and transparency and disclosure (WP5). MENTAL.AI@CaeSAR relies on an international, interdisciplinary consortium bringing together researchers in computer science, neuroscience, psychiatry, linguistics, cognitive science, and law. By combining these areas of expertise, the project aims to lay the foundation for a new generation of digital tools capable of assisting clinicians in decision-making and improving patient care. In the long term, this work could reshape practices in mental and brain health, supporting earlier diagnoses, precision treatments, and patient-centered monitoring that protects essential neuro-rights.

Start and end dates: 12/2025 - 12/2029
Principal Investigator: Gaël Dias, GREYC UMR 6072



$*$ *CaeSAR: Caen, Stratégie pour l'Accélération en Recherche*
$**$ *A2MIMO: Artificial Agent for Mind Monitoring*

Topics of interest for the RaPID-6@MENTAL.ai

The topics of interest for the workshop session included but were *not* limited to:

- Guidelines, methods and protocols for (remote) data collection and/or annotation (schemas, tools)
- Infrastructure for the domain: building, adapting and sharing of linguistic resources, data sets and tools
- Acquisition and combination of novel data samples; including digital biomarkers, continuous streaming, monitoring and aggregation of measurements; as well as self-reported behavioral and/or physiological and activity data
- Addressing the challenges of representation, including dealing with data sparsity and dimensionality issues, feature combination from different sources and modalities
- Domain adaptation of NLP/AI tools
- Acoustic/phonetic/phonologic, syntactic, semantic, pragmatic and discourse analysis of data; including modeling of perception (e.g. eye-movement measures of reading) and production processes (e.g. recording of the writing process by means of digital pens, keystroke logging etc.); use of gestures accompanying speech and non-linguistic behavior
- Use of wearable, vision, and ambient sensors or their fusion for detection of cognitive disabilities or decline
- (Novel) Modeling and deep / machine learning approaches for early diagnostics, (severity) prediction, monitoring, classification, such as:
  - multimodal learning
  - large pre-trained Transformer language models [LLMs]
  - explainable and interpretable AI models
- Evaluation of the significance of features for screening and diagnostics
- Evaluation of tools, systems, components, metrics, applications and technologies including methodologies making use of NLP/AI; e.g. for predicting clinical scores from (linguistic and/or digital) features
- Digital platforms/technologies for cognitive assessment and brain training
- Evaluation, comparison and critical assessment of resources
- Involvement of medical/clinical professionals and patients
- Ethical, gender bias, legal and safety questions in research with human data in the domain, and how they can be handled
- Deployment, assessment platforms and services as well as innovative mining approaches that can be translated to practical/clinical applications
- Experiences, lessons learned and the future of NLP/AI in the area

Organizing Committee

- Dimitrios Kokkinakis, University of Gothenburg, Sweden *(Workshop's chair)*
- Charalambos Themistocleous, University of Oslo, Norway *(Workshop's co-chair)*
- Gaël Dias, University of Caen Normandie, France
- Kathleen C. Fraser, University of Ottawa, Canada
- Fredrik Öhman, University of Gothenburg, Sweden
- Sebastião Pais, University of Beira Interior, Portugal

Scientific Programme Committee (in alphabetic order)

- Chiara Barattieri di San Pietro, Scuola Universitaria Superiore IUSS di Pavia, Italy
- Patrice Bellot, Aix-Marseille University, France
- Visar Berisha, Arizona State University, USA
- Eric Bui, Caen University Hospital, France
- Gaël Chételat, INSERM, France
- Sunghye Cho, University of Pennsylvania, USA
- Corinne Fredouille, Avignon University, France
- Valantis Fyndanis, University of Technology, Cyprus and University of Oslo, Norway
- Gloria Gagliardi, University of Bologna, Italy
- Natalia Grabar, CNRS, France
- Martínez-Nicolás Israel, Universidad de Salamanca, Spain
- Sandra Anna Just, UiT – The Arctic University, Norway
- Alexandra König, Ki-Elements, Germany
- Klervi Le Dortz, University of Caen Normandie, France
- Alice Lee, University College Cork, Ireland
- Mark Lee, University of Birmingham, UK
- Kristina Lundholm Fors, University of Lund, Sweden
- Patricia Martín-Rodilla, Spanish National Research Council, Spain
- Fabrice Maurel, University of Caen Normandie, France
- Jérémie Pantin, University of Caen Normandie, France
- Javier Parapar, University of La Coruña, Spain
- Alexandre Pauchet, INSA Rouen Normandy, France

- Emily Prud'hommeaux, Boston College, USA
- Masoud Rouhizadeh, University of Florida, USA
- Lina Rydén, University of Gothenburg, Sweden
- Sriparna Saha, Indian Institute of Technology Patna, India
- Johan Skoog, University of Gothenburg, Sweden
- Athanasios Tsanas, University of Edinburgh, UK
- Spyridoula Varlokosta, National and Kapodistrian University of Athens, Greece
- Yaru Wu, University of Caen Normandie, France
- Yasunori Yamada, Boston Medical Science Co., USA

Table of Contents

RaPID-6@MENTAL.ai Workshop Program

Tuesday, May 12, 2026

14:00–16:00 **Session A**
Chair: Dimitrios Kokkinakis

14:00–14:05 **Welcome and Introduction**

14:05–14:50 **Keynote Speaker 1: Brian MacWhinney, Teresa Heinz Professor of Cognitive Psychology, Carnegie Mellon University, USA: "TalkBank resources for studying functional communication in language disorders" (online)**

14:55–16:00 **Oral Session**

14:55–15:15 *Multilingual Cognitive Impairment Detection in the Era of Foundation Models*
Damar Hoogland, Boshko Koloski, Jaya Caporusso, Tine Kolenik, Senja Pollak, Christina Manouilidou and Matthew Purver

15:15–15:35 *The Icelandic Language Biobank: Data Collection through a Clinical Analysis Platform*
Iris Nowenstein, Naizeth Núñez Macías, Gunnar Thor Örnólfsson, Stefán Ólafsson, Bryndís Bergþórsdóttir, Iðunn Kristínardóttir and Hinrik Hafsteinsson

15:35–15:55 **Invited Speaker: Professor Gaël Dias, Université de Caen Normandie, France: "MENTAL.ai - Artificial Intelligence for Mental and Brain Health"**

15:55–16:00 **Questions or comments to the sessions's presenters**

16:00–16:30 ***Afternoon Coffee Break***

Tuesday, May 12, 2026 (continued)

16:00–18:15 **Session B: Poster and Oral Sessions**
Chair: Charalambos Themistocleous

16:00–18:20 **Poster Session** (Note: The poster session will begin during the coffee break. Poster presenters are expected to be available at their posters until 17:00, while the posters will remain on display until the end of the workshop)

Disfluencies and ASR Performance on Swedish Spontaneous Speech from the 'Trip to Stockholm' Discourse Narrative Task
Dimitrios Kokkinakis, Herbert Lange and Ricardo Muñoz Sánchez

ALBA: An Automated Framework for Benchmarking Clinical Language Biomarkers against Standardized Corpora
Charalambos Themistocleous and Brielle C. Stark

On Automatic Detection of Cognitive Decline
Fabio Tamburini

Benchmarking NLP-supported Language Sample Analysis for Swiss Children's Speech
Anja Ryser, Yingqiang Gao and Sarah Ebling

Resource-Efficient LLMs for Depression Symptoms Screening: Performance and Limitations in Zero Shot Setting
Muhammad Rizwan and Jure Demšar

CNSocialDepress: A Chinese Social Media Dataset for Depression Risk Detection and Structured Analysis
Jinyuan Xu, Tian Lan, Xintao Yu, Xue He, Hezhi Zhang, Ying Wang, Mathieu Valette, Pierre Magistry and Lei Li

Depression Detection in Modern Greek
Vivian Stamou, George Mikros, George Markopoulos and Spyridoula Varlokosta

Profiling Psychopathic Behavior Using Machine Learning
Avi Treistman, Tehilla David, Sivan Levi and Dror Mughaz

Tuesday, May 12, 2026 (continued)

17:00–17:40 **Keynote Speaker 2: Sunny X. Tang, M.D., Assistant Professor, Department of Psychiatry, Feinstein Institutes for Medical Research / Zucker Hillside Hospital, Glen Oaks, NY, USA: "Speech and Language Markers for Clinical Applications in Psychiatric Disorders" (online)**

17:40–18:20 **Oral Session**

17:40–17:55 *Developing Annotation Guidelines for CSAM Prevention Interventions: Psychosocial Risk and Protective Factors Grounded in Research and Clinical Practice*
Vera Czehmann, Christine Hovhannisyan, Lena Elisabeth Hoffmann, Paula Busch, Ibrahim Baroud, Sebastian Möller, Roland Roller, Hannes Gieseler and Lisa Raithel

17:55–18:10 *Automatic Detection of Direct and Self-Repetitions in Naturalistic Speech Recordings of French- and Dutch-Speaking Autistic Children*
Federica Beccaria, Marie Kolenberg, Pierre Labendzki, Inge Zink and Mikhail Kissine

18:10–18:15 **Questions or comments to the sessions's presenters**

18:15–18:20 **Conclusions**