Editorial: The Third Workshop on Explainable Artificial Intelligence for the medical domain - EXPLIMED@IJCAI-ECAI2026

Gabriella Casalino[1], Giovanna Castellano[1], Katarzyna Kaczmarek-Majer[2,3], Alberto Gaetano Valerio[1] and Gianluca Zaza[1]

[1]*Computer Science Department, University of Bari Aldo Moro Bari, Italy*

[2]*Systems Research Institute Polish Academy of Sciences – Warsaw, Poland*

[3]*University of Ostrava, Institute for Research and Applications of Fuzzy Modeling, NSC IT4Innovations – Ostrava, Czech Republic*

Abstract

The Third Workshop on Explainable Artificial Intelligence for the Medical Domain (EXPLIMED 2026) was held in conjunction with the 35th International Joint Conference on Artificial Intelligence (IJCAI-ECAI 2026), continuing the workshop's mission of advancing research on transparent, trustworthy, and human-centered Artificial Intelligence for healthcare. Building upon the success of the previous editions, EXPLIMED brought together researchers from academia and industry to discuss recent methodological advances, emerging applications, and open challenges in Explainable Artificial Intelligence (XAI) for the medical domain. Following a rigorous peer-review process, 10 high-quality papers were accepted from 22 submissions, covering a broad spectrum of topics including medical imaging, physiological signal analysis, genomics, knowledge-guided reasoning, conversational AI, and human-centered explainability. The workshop provided an interdisciplinary forum for exchanging ideas on novel explainability methodologies, trustworthy AI, clinical decision support, and the evaluation of explanation quality, fostering collaborations across the Artificial Intelligence and healthcare communities.

Keywords

Explainable Artificial Intelligence, Trustworthy Artificial Intelligence, Healthcare AI, Clinical Decision Support, Medical Data Analysis

1. Introduction

Artificial Intelligence (AI) is rapidly transforming modern healthcare by supporting clinicians in diagnosis, prognosis, treatment planning, patient monitoring, and personalized medicine. The increasing availability of multimodal biomedical data, together with recent advances in machine learning, deep learning, and foundation models, has enabled the development of highly accurate predictive systems across a wide range of medical applications [1]. Nevertheless,

EXPLIMED 2026 - Third Workshop on Explainable Artificial Intelligence for the medical domain - 15-17 August 2026, Bremen, Germany

*Corresponding author.

gabriella.casalino@uniba.it (G. Casalino); giovanna.castellano@uniba.it (G. Castellano); K.Kaczmarek@ibspan.waw.pl (K. Kaczmarek-Majer); a.valerio31@phd.uniba.it (A. G. Valerio); gianluca.zaza@uniba.it (G. Zaza)

000-0003-0713-2260 (G. Casalino); 0000-0002-6489-8628 (G. Castellano); 0000-0003-0422-9366 (K. Kaczmarek-Majer); 0009-0008-6101-965X (A. G. Valerio); 0000-0003-3272-9739 (G. Zaza)

© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR-WS.org/Vol-4339/preface.pdf

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

published 2026-11-26

achieving high predictive performance alone is no longer sufficient for the deployment of AI systems in real clinical practice. Healthcare professionals must be able to understand, verify, and appropriately trust AI-assisted recommendations before integrating them into clinical decision-making processes [2]. Consequently, explainability has become a fundamental requirement for ensuring the safe, transparent, and responsible adoption of AI in healthcare [3, 4].

While Explainable Artificial Intelligence (XAI) was initially introduced to address the opacity of complex machine learning models, recent research has progressively shifted towards the broader concept of Trustworthy AI [5, 6]. In high-stakes domains such as medicine, explainability represents only one component of trustworthy intelligent systems, which must also satisfy requirements related to robustness, reliability, fairness, accountability, uncertainty awareness, and regulatory compliance. Recent regulatory initiatives, including the European AI Act, further emphasize the importance of transparency and human oversight in AI systems intended for healthcare applications, reinforcing the need for explainability methods that are both technically sound and clinically meaningful [7].

At the same time, the rapid evolution of foundation models and Generative Artificial Intelligence has introduced new opportunities as well as new challenges for medical AI. Vision-language models, multimodal foundation models, and large language models are increasingly being incorporated into clinical workflows, enabling richer reasoning capabilities, natural-language interaction, and more intuitive explanations for end users. However, these models also introduce additional concerns regarding hallucinations, factual consistency, uncertainty estimation, and the transparency of complex reasoning processes. As a result, explainability is evolving from the interpretation of isolated model predictions towards the explanation of increasingly sophisticated multimodal and interactive AI systems [8].

Another important research trend concerns the transition from algorithm-centered to human-centered explainability. Rather than evaluating explanations exclusively through technical metrics, recent studies increasingly recognize the importance of assessing their usefulness for clinicians and other stakeholders [9]. Effective explanations should support clinical reasoning, fit naturally within existing healthcare workflows, communicate uncertainty, and improve users' confidence without increasing cognitive burden. Consequently, explanation quality is now being evaluated not only in terms of fidelity and faithfulness but also with respect to clinical relevance, usability, actionability, and human-AI collaboration [10].

The growing adoption of XAI has also highlighted the need for rigorous methodologies to evaluate explanation quality. While early studies often relied on qualitative examples or subjective visual inspection, recent research increasingly advocates for systematic evaluation protocols capable of measuring multiple dimensions of explainability [11]. Beyond traditional properties such as fidelity and faithfulness, explanations are now assessed with respect to stability, robustness, completeness, plausibility, actionability, and clinical usefulness. Moreover, there is growing recognition that no single evaluation metric is sufficient, motivating the development of comprehensive benchmarking frameworks that combine quantitative measures with human-centered assessments involving clinicians and domain experts. Establishing reliable evaluation methodologies is therefore becoming a fundamental prerequisite for comparing explainability techniques and facilitating their adoption in real-world clinical settings [12].

Alongside these developments, the methodological landscape of XAI continues to diversify. Beyond traditional feature-attribution techniques, recent research increasingly explores concept-

based explanations [13], inherently interpretable models [14], neuro-symbolic approaches [15], knowledge-guided reasoning [16], graph-based explainability [17, 18], conversational interfaces [19], and uncertainty-aware explanation frameworks [**?**]. These complementary paradigms aim to produce explanations that are more robust, intuitive, and aligned with human reasoning, while maintaining the predictive capabilities of modern AI models [20].

Against this rapidly evolving background, the Third Workshop on Explainable Artificial Intelligence for the Medical Domain (EXPLIMED 2026), held in conjunction with the 35$^{\text{th}}$ International Joint Conference on Artificial Intelligence (IJCAI-ECAI 2026), continues to provide an international forum for researchers working at the intersection of Artificial Intelligence, explainability, and healthcare. The workshop promotes interdisciplinary dialogue among researchers, clinicians, and practitioners interested in developing transparent, trustworthy, and human-centered AI systems for medicine. The accepted contributions span a broad range of topics, including medical imaging, physiological signal analysis, genomics, computational pathology, conversational AI, knowledge-guided reasoning, and explainability evaluation, reflecting both the diversity of current XAI methodologies and the growing maturity of the field.

2. Keynote

The workshop opened with an invited keynote delivered by Prof. Jerzy Stefanowski (Poznan University of Technology, Poland), entitled *Challenges in Evaluating Explanation Methods for Static and Evolving Models*.

The keynote focused on one of the most pressing challenges in Explainable Artificial Intelligence (XAI): the need for rigorous and meaningful evaluation of explanation methods. While the rapid growth of XAI has led to the development of a large number of explanation techniques, considerably less attention has been devoted to understanding how their effectiveness should be assessed and how explanations can be validated in real-world scenarios. Prof. Stefanowski highlighted the importance of complementing conventional quantitative metrics with human-centred evaluation protocols capable of assessing the usefulness, interpretability, and practical impact of explanations.

The presentation illustrated how explainability can support the identification and mitigation of undesirable biases in machine learning systems, demonstrating that XAI can play an active role not only in interpreting model predictions but also in improving the fairness and reliability of intelligent systems, see e.g., [21]. The keynote also emphasized the current lack of benchmark datasets and standardized evaluation procedures for explanations, arguing that closer collaboration with domain experts is essential for establishing reliable evaluation methodologies and strengthening users' trust in AI-assisted decision making.

A significant part of the talk was devoted to the emerging problem of explainability in dynamic learning environments. Unlike traditional static settings, models operating on evolving data streams require explanations that remain consistent and informative as both data distributions and predictive models change over time. The keynote discussed recent research directions aimed at extending explanation methods to these adaptive scenarios and highlighted the need for novel evaluation frameworks capable of jointly assessing the temporal evolution of data, models, and explanations.

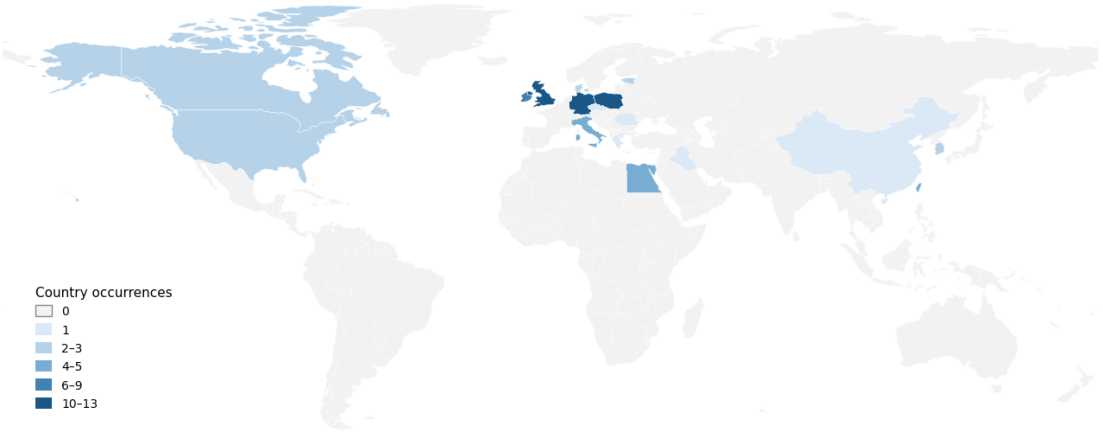


Figure 1: Geographical distribution of the countries represented in the authors' affiliations of the workshop submissions. Color intensity indicates the frequency of occurrence of each country.

The keynote attracted significant interest from the workshop participants and stimulated an engaging discussion on the future directions of XAI research. In particular, it reinforced the importance of shifting the focus from the development of increasingly sophisticated explanation methods towards establishing robust evaluation methodologies and promoting closer collaboration between AI researchers, domain experts, and end users.

3. Workshop's Contributions

The third edition of the EXPLIMED workshop received 22 submissions, from which 10 papers were accepted after a rigorous peer-review process. The submissions involved researchers from 17 countries, namely Romania, Iraq, Ireland, Denmark, Czechia, Italy, Greece, the United Kingdom, Egypt, Germany, Taiwan, South Korea, China, Estonia, Canada, Poland, and the United States. As shown in Figure 1, the workshop attracted contributions from a geographically diverse research community spanning Europe, Asia, North America, and Africa. This broad international participation confirms the growing interest in Explainable Artificial Intelligence for healthcare and highlights the relevance of the topic across different research communities.

As a joint workshop of IJCAI-ECAI 2026, EXPLIMED provided an international forum for researchers working at the intersection of Artificial Intelligence, explainability, and medicine. The diversity of participating countries and research perspectives further emphasizes the global effort towards developing transparent, trustworthy, and clinically meaningful AI systems for healthcare.

To obtain an overview of the main research themes emerging from the accepted contributions, we conducted a content analysis of the paper abstracts. As a first exploratory step, we generated the word cloud shown in Figure 2, which summarizes the most frequently occurring terms across the accepted papers. The prominence of terms such as clinical, XAI, attribution, explanation, model, framework, and classification reflects the workshop's strong emphasis on developing

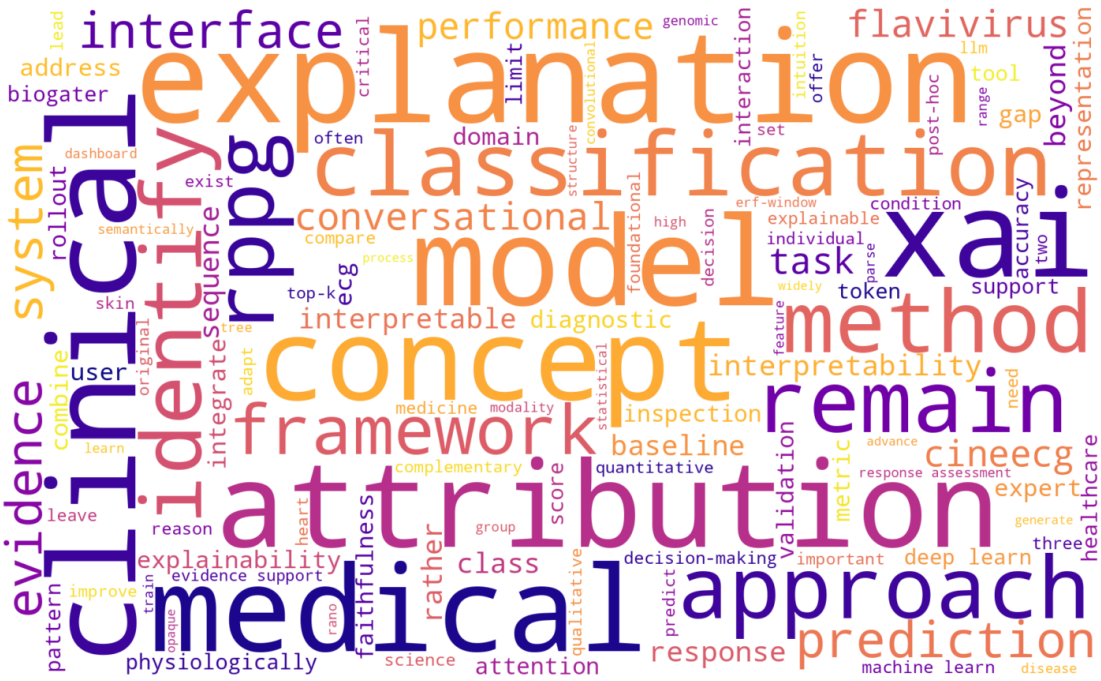


Figure 2: Wordcloud showcasing the most frequently occurring terms in the abstract of the workshop documents.

explainable AI methods for clinically relevant applications. Beyond these dominant concepts, the presence of terms such as evaluation, validation, interpretability, interface, conversational, and support highlights the growing attention devoted to assessing explanation quality and designing explainability solutions that effectively assist end users. At the same time, more specialized terms, including physiologically, attention, flavivirus, token, CineECG, diagnostic, and sequence, reveal the diversity of application domains, data types, and methodological paradigms explored by the accepted contributions, ranging from medical imaging and physiological monitoring to genomics, multimodal learning, and human-centered interaction. The resulting word cloud therefore captures both the common research focus on trustworthy clinical AI and the broad spectrum of explainability challenges addressed throughout the workshop.

To further characterize the accepted papers, we performed a structured analysis from five complementary perspectives: primary data modality, primary research objective, primary explainability approach, primary AI methodology, and primary clinical domain. For each perspective, every contribution was assigned to a single primary category according to its predominant focus. The resulting distributions are summarized in Figure 3. Together, these complementary analyses provide a comprehensive overview of the current research landscape in Explainable Artificial Intelligence for the medical domain, illustrating the diversity of data modalities, methodological approaches, research objectives, and clinical applications represented within the workshop. They also highlight the interdisciplinary nature of contemporary XAI research, where advances in machine learning, human-centered AI, knowledge representation, and clinical decision support increasingly converge.

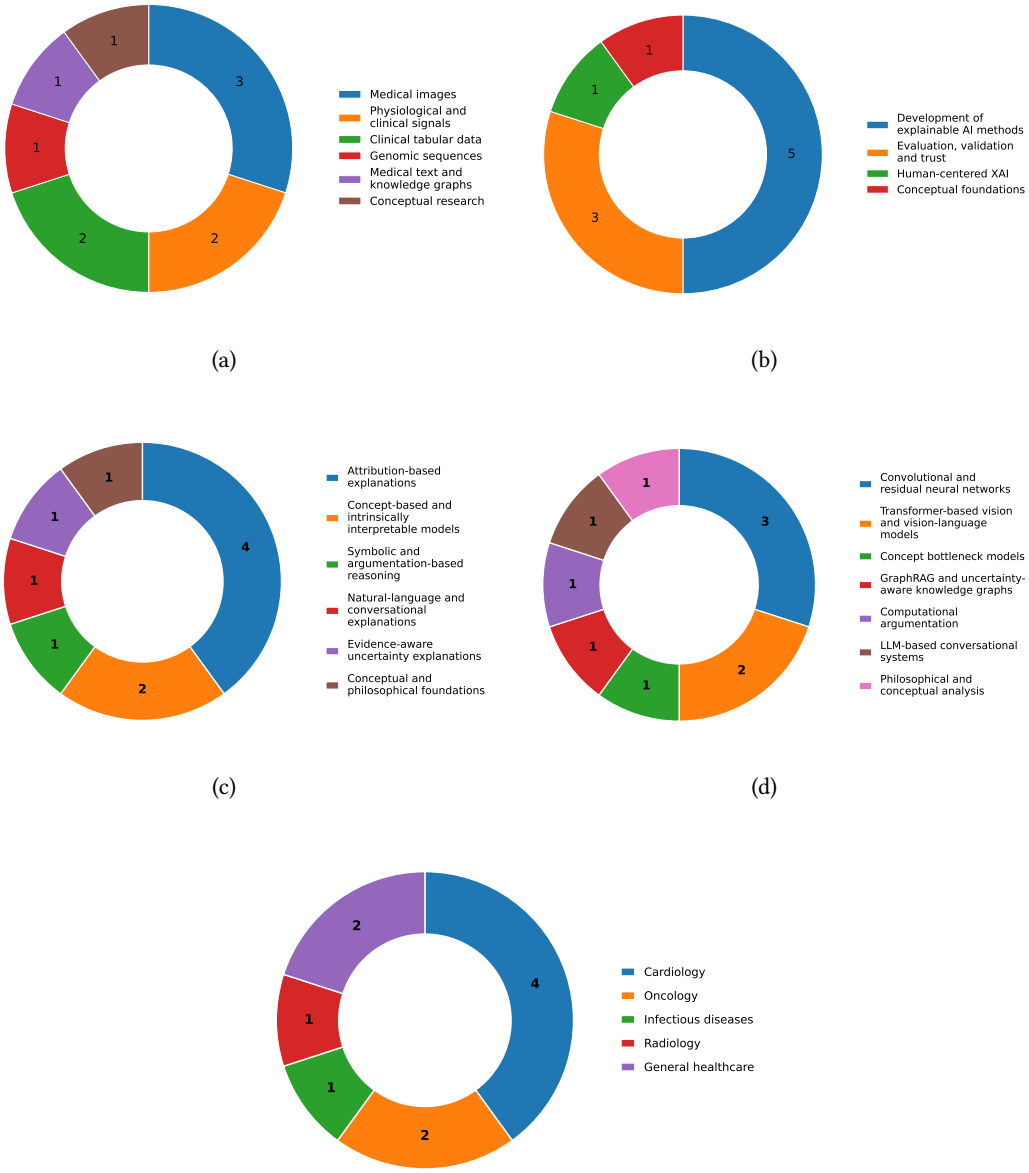


Figure 3: Overview of the accepted contributions according to five complementary dimensions: **(a)** primary data modality; **(b)** primary research objective; **(c)** primary explainability approach; **(d)** primary AI methodology; and **(e)** primary clinical domain. Each contribution was assigned to a single primary category within each dimension according to its predominant focus.

Figure 3(a) summarizes the primary data modalities addressed by the accepted contributions. Although medical imaging represents the largest group, the workshop also includes studies on physiological and clinical signals, structured clinical data, genomic sequences, and medical text. This diversity highlights that explainability is no longer limited to image analysis but is becoming an essential requirement across heterogeneous biomedical data sources, each posing distinct interpretability challenges.

Figure 3(b) categorizes the contributions according to their primary research objective. A substantial portion of the papers focuses on the development of novel explainable AI methodologies, while several contributions investigate the evaluation, validation, and trustworthiness of existing explanation techniques. Human-centred explainability and conceptual foundations are also represented, reflecting the growing awareness that effective XAI requires not only methodological innovation but also rigorous evaluation protocols and a deeper understanding of how explanations should support human reasoning and decision-making.

Figure 3(c) presents the distribution of the main explainability paradigms adopted in the workshop. Attribution-based explanations remain the most frequently explored approach, confirming their central role in interpreting complex predictive models. At the same time, the accepted contributions cover a broad spectrum of complementary paradigms, including concept-based and inherently interpretable models, symbolic reasoning, conversational explanations, uncertainty-aware explanation frameworks, and conceptual analyses. This variety illustrates the ongoing evolution of XAI from a predominantly post-hoc perspective towards richer and more diverse forms of explainability tailored to different application scenarios.

The methodological landscape is further illustrated in Figure 3(d). The contributions encompass a wide range of Artificial Intelligence techniques, including convolutional and transformer-based neural networks, concept bottleneck models, knowledge graph-based systems, computational argumentation, Large Language Models, and symbolic approaches. This methodological diversity demonstrates that explainability is increasingly regarded as a cross-cutting research objective rather than a property of a specific family of AI models, motivating the development of dedicated XAI solutions across heterogeneous computational paradigms.

Finally, Figure 3(e) shows the distribution of the addressed clinical domains. Cardiology represents the most represented area, followed by oncology, while additional contributions target infectious diseases, radiology, and general healthcare scenarios. Although the accepted papers cover a broad range of clinical applications, the predominance of cardiovascular and oncological studies reflects the strong demand for transparent AI systems in high-impact medical specialties where explainability can directly support clinical decision-making and increase users' confidence in AI-assisted healthcare.

Taken together, these five complementary perspectives confirm that research on Explainable Artificial Intelligence for healthcare is rapidly expanding beyond individual algorithms or application domains. The accepted contributions demonstrate that explainability has become a unifying research theme spanning different data sources, methodological paradigms, evaluation strategies, and clinical contexts, ultimately contributing to the development of more transparent, trustworthy, and clinically useful AI systems.

3.1. Session 1

The first technical session featured five contributions covering a broad range of topics in Explainable Artificial Intelligence for healthcare. A brief overview of each paper is provided below.

The paper *Scientific Explanations in Health Sciences: Causality, Trust, and Epistemic Adequacy [22]* examines the conceptual foundations of Explainable Artificial Intelligence (XAI) in healthcare by connecting medical AI with the philosophy of science. The paper argues that effective medical explanations should go beyond improving model transparency and should instead support clinical reasoning, informed decision-making, and justified trust. To this end, it discusses three key dimensions of medical explainability—causality, trust, and explanatory adequacy—and proposes a set of philosophically grounded design principles for developing XAI systems that better address the needs of clinicians, patients, and other stakeholders. The work provides a valuable conceptual perspective on the development of trustworthy and clinically meaningful medical AI.

The paper *Evaluating the Role of Data and Explanation Representations in Clinical XAI for Conduction Disorder Classification [23]* investigates how both data representation and explanation format influence the effectiveness of Explainable Artificial Intelligence in clinical decision support. Using the task of conduction disorder classification from electrocardiograms, the authors compare conventional 12-lead ECGs with the 3D CineECG representation, together with two complementary explanation techniques: attribution heatmaps and rule-based validity corridors. Beyond assessing classification performance, the study includes an expert evaluation involving clinicians to analyze the interpretability, clinical relevance, and usefulness of the different explanation formats. The results show that the standard 12-lead ECG remains the preferred representation, while CineECG provides complementary information, highlighting the importance of designing clinical XAI systems that combine familiar diagnostic views with multiple forms of explanation.

The paper *Explaining RhythmFormer: A Systematic XAI Analysis of Periodic Sparse Attention for Remote Photoplethysmography [24]* presents a comprehensive framework for evaluating explainability in transformer-based remote photoplethysmography models. Focusing on the RhythmFormer architecture, the authors systematically compare multiple attention-based attribution methods, including raw attention, attention rollout, attention flow, and Beyond Intuition. To quantitatively assess explanation quality, they introduce two complementary evaluation metrics: skin coverage, measuring the alignment of explanations with physiologically relevant facial regions, and an adaptation of the SaCo faithfulness metric to regression tasks. The experimental analysis highlights the strengths and limitations of different explanation techniques, demonstrating that Beyond Intuition provides the most faithful and physiologically meaningful explanations, while emphasizing the importance of rigorous quantitative evaluation for trustworthy AI in contactless physiological monitoring.

The paper *U-GraphRAG: Evidence-Support Estimation in a Medical GraphRAG System [25]* introduces U-GraphRAG, an uncertainty-aware GraphRAG framework designed to estimate the degree of evidence supporting responses generated for medical question answering. The proposed methodology constructs a hierarchical uncertainty-aware knowledge graph by extracting relational triples from medical documents using multiple large language models and enriching

them with confidence indicators, source reliability, recurrence-based evidence, and query relevance. These heterogeneous uncertainty signals are propagated through fuzzy aggregation operators to compute an interpretable evidence-support score accompanying each generated response. Experimental results on synthetic medical question-answering scenarios demonstrate the ability of the proposed framework to distinguish well-supported from weakly supported answers, highlighting the potential of uncertainty-aware GraphRAG systems for improving transparency and trustworthiness in AI-assisted clinical decision support.

The paper *TRACE: A Concept Bottleneck Model for Longitudinal 3D Glioblastoma Response Assessment [26]* presents TRACE, a concept bottleneck model for interpretable longitudinal assessment of glioblastoma treatment response from multimodal MRI. The proposed framework follows the RANO 2.0 clinical guidelines by predicting clinically meaningful intermediate concepts, such as tumor measurements and response indicators, before inferring the final response category. By explicitly encoding the clinical reasoning process through a structured concept graph, the model enables transparent decision-making and allows clinicians to inspect and potentially correct intermediate concepts. Experimental results on the publicly available LUMIERE dataset demonstrate that TRACE achieves competitive predictive performance while improving interpretability compared with conventional deep learning approaches, highlighting the potential of concept-based models for trustworthy AI-assisted neuro-oncology.

3.2. Session 2

The second session featured five additional contributions that further expanded the range of research topics presented at the workshop. Complementing the themes discussed in the first session, these papers explored novel methodologies, evaluation strategies, and application scenarios, providing additional insights into the rapidly evolving landscape of Explainable Artificial Intelligence for the medical domain. A brief overview of each contribution is presented below.

The paper *Evaluating Conversational Explainability in Healthcare: Comparing the Explainability Assistant and a Traditional XAI Dashboard [27]* investigates whether conversational interfaces improve the accessibility of explainable AI for healthcare professionals compared with traditional dashboard-based XAI tools. The authors conduct a within-subjects user study comparing the Explainability Assistant, an LLM-based conversational XAI system, with a conventional dashboard, while keeping the underlying prediction model (a Decision Tree for heart disease classification) and explanation methods (SHAP, DiCE, and what-if analysis) identical across both interfaces. The study, involving healthcare practitioners and students, combines quantitative and qualitative evaluations, showing that conversational interfaces facilitate usability and understanding, whereas dashboards provide better visibility of global and local explanations. Rather than identifying a clear winner, the findings suggest that the two interaction paradigms are complementary, supporting the development of integrated clinical XAI tools that combine conversational and visual explanations.

The paper *Explaining BiomedCLIP with Weighted Banzhaf Interactions Supported by Tree-Gram Parsing [28]* addresses the challenge of generating faithful and clinically meaningful explanations for vision-language models in medical imaging. The authors propose ParseFIxLIP, a linguistically informed extension of the FIxLIP framework that leverages dependency-tree

parsing to group fragmented subword tokens into semantically coherent medical concepts before computing game-theoretic explanations. Extensive quantitative and qualitative experiments on the ROCOv2 dataset demonstrate that the proposed semantic grouping strategy improves explanation faithfulness, robustness, and interpretability compared with token-level baselines, while mitigating the curse of dimensionality inherent in long clinical reports. Moreover, the proposed framework provides valuable insights into the behavior and limitations of BiomedCLIP, revealing issues such as shortcut learning from visual annotations and sensitivity to spatial transformations, thereby contributing to more trustworthy and clinically relevant explainable AI for multimodal medical foundation models.

The paper *Interpretable Flavivirus Classification under Low-Coverage Conditions via ERF-Aware k-mer Attribution [29]* presents an architecture-aware explainability framework for deep learning-based flavivirus classification from genomic sequences. Building upon the DiCNN-UniK convolutional neural network, the proposed approach traces high-activation convolutional responses back to the corresponding k-mer windows within the model's effective receptive field, providing sequence-level explanations aligned with the model's internal decision-making process. The framework further evaluates the robustness and faithfulness of the identified genomic signatures through simulated nucleotide ambiguity and occlusion-based perturbation analyses. Experimental results demonstrate that the proposed attribution method identifies discriminative and biologically plausible sequence patterns while maintaining robust explanations under low-coverage sequencing conditions. This work contributes to the development of transparent and trustworthy AI methods for genomic surveillance and pathogen monitoring.

The paper *HistoDyn-BioGater:: Biologically Interpretable Dynamic Gating Architecture for Computational Pathology [30]* presents HistoDyn-BioGater, a lightweight and biologically interpretable vision architecture for computational pathology. Instead of relying on conventional convolutional or transformer blocks, the proposed framework employs a dynamically gated collection of biologically inspired operations, including stain normalization, contrast enhancement, edge detection, and morphological transformations, enabling the model to selectively activate the most informative processing pathways. To further support interpretability, the authors introduce the Gated Activation Diversity Score (GADS), a novel metric for quantifying gate specialization and operational diversity. Experimental results on multiple histopathology and endoscopic imaging datasets demonstrate that the proposed architecture improves classification performance, calibration, and computational efficiency while providing transparent, domain-aligned explanations, highlighting its potential for trustworthy and scalable AI-assisted computational pathology.

The paper *ARGO: An Argumentative Framework for Guided Medical Outcomes [31]* presents a symbolic explainable AI framework for medical diagnosis based on Computational Argumentation. Rather than relying on black-box predictive models, the proposed approach learns interpretable diagnostic theories directly from clinical data and explicitly represents diagnostic uncertainty by identifying argumentative dilemmas when the available evidence does not support a single reliable conclusion. To enhance usability in clinical practice, the framework integrates a prompt-constrained Large Language Model that translates formal argumentation traces into clear, human-readable diagnostic explanations while preserving the underlying logical reasoning. Experimental results on multiple medical datasets demonstrate the framework's ability to produce accurate, transparent, and trustworthy diagnostic support, highlighting the

potential of argumentation-based reasoning for explainable clinical decision support systems.

4. Organizing Committee



Gabriella Casalino is an Assistant Professor (Tenure Track) at the Computational Intelligence Laboratory (CILab), Department of Informatics, University of Bari Aldo Moro, Italy. Her research focuses on Computational Intelligence and Explainable Artificial Intelligence (XAI) for eHealth, with particular emphasis on trustworthy and human-centered AI, fuzzy systems, data stream mining, and machine learning under limited supervision. She is also a member of the PhD Board in Digital Innovation and e-Health at the University of Bari Aldo Moro and the Interdepartmental Center for Telemedicine (CITEL) of University of Bari. She is a member of the IEEE Task Force on Explainable Fuzzy Systems, and currently serves as Secretary of EUSFLAT – the European Society for Fuzzy Logic and Technology. Gabriella is actively involved in the international AI community, organizing workshops and special sessions at leading conferences, including IJCAI/ECAI and IEEE WCCI. She serves as Associate Editor for IEEE Transactions on Computational Social Systems, Scientific Reports, and Soft Computing, and has been Guest Editor for special issues in IEEE Transactions on Computational Social Systems and Information Systems Frontiers. She is a Senior Member of IEEE, has participated in numerous national and international research projects, and has co-authored more than 90 scientific publications, including work recognized with the Best Paper Award at FUZZ-IEEE 2022.



Giovanna Castellano is a Full Professor at the Department of Computer Science, University of Bari Aldo Moro, where she coordinates the Computational Intelligence Laboratory (CILab). Her research interests are in the area of Computational Intelligence and Computer Vision. She has been responsible for the local unit of several research projects and is currently the Principal Investigator of the WP 6.4 "Understandability of AI systems" in the NRRP "FAIR - Future Artificial Intelligence Research" project, Spoke 6 - Symbiotic AI. She is an Associate Editor of several international journals. She has been a Guest Editor of special issues and participated in organizing scientific events. She is a reviewer for several international journals published by leading publishers, including Elsevier, IEEE, and Springer, and a member of the program committee of several international conferences. She is a member of the IEEE Society, EUSFLAT Society, INDAM-GNCS Society, IAPR Technical Committee 19 (Computer Vision for Cultural Heritage Applications), CINI-AIIS laboratory, CINI-BIG DATA laboratory, CITEL telemedicine research center, GRIN, MIR laboratories. She is also a member of the IEEE CIS

Task Force on Explainable Fuzzy Systems.



Katarzyna Kaczmarek-Majer is an Associate Professor at the Systems Research Institute of the Polish Academy of Sciences. Katarzyna serves as the Principal Investigator for the "Explainable Artificial Intelligence for Monitoring Acoustic Features Extracted from Speech" project (ExplainMe), funded by the European Regional Development Fund. Additionally, she contributes to the "Research of Excellence on Digital Technologies and Wellbeing" project at the Institute for Research and Applications of Fuzzy Modeling, University of Ostrava, Czech Republic. Katarzyna's expertise includes computational intelligence, explainable and trustworthy AI, granular computing, and data stream analysis, with a particular focus on applications in medicine and healthcare. She has co-authored over 50 scientific publications, some of which have been recognised with prestigious awards, such as the Best Paper Award at FUZZ-IEEE 2022 in Padova, Italy, for her work titled "Confidence Path Regularization for Handling Label Uncertainty in Semi-Supervised Learning: A Use Case in Bipolar Disorder Monitoring." She serves on the scientific committees of numerous conferences and is a reviewer for several academic journals and international conferences. Katarzyna is also the Vice-President for the European Society for Fuzzy Logic and Technology and the President of the eHealth Section of the Polish Information Processing Society.



Alberto Gaetano Valerio is a Ph.D. candidate in Digital Innovation and e-Health at the University of Bari Aldo Moro, Italy, where he conducts research within the Computational Intelligence Laboratory (CILab). His research focuses on trustworthy artificial intelligence for healthcare, with particular interests in explainable AI, graph neural networks, computer vision, multimodal learning, and large language models for clinical decision support. His current work investigates interpretable deep learning methods for medical imaging and speech analysis, with applications including brain tumor segmentation, neurodegenerative diseases, mental health assessment, and drug–drug interaction analysis. He is also interested in developing evaluation methodologies for AI explanations in clinical settings, combining quantitative metrics with human-centered assessment. Alberto has published research in international venues in artificial intelligence and medical imaging and serves as a reviewer for several journals and conferences in these fields. His research has also included a visiting research period at the Poznań University of Medical Sciences, focusing on graph-based learning methods for mental health assessment.



Gianluca Zaza is an Assistant Professor at the Department of Computer Science, University of Bari Aldo Moro (Italy), and a member of the Computational Intelligence Laboratory (CILab). His research focuses on Explainable Artificial Intelligence, trustworthy and human-centred AI, fuzzy systems, computational intelligence, and AI for healthcare, with particular emphasis on the understandability of intelligent systems. He has been actively involved in the activities of the NRRP project FAIR – Future Artificial Intelligence Research, Spoke 6 Symbiotic AI, contributing to research on explainable and human-centred AI methodologies.

He was the Principal Investigator of the GNCS-funded research project Computational Models based on Fuzzy Logic for eXplainable Artificial Intelligence and is a member of the Interdepartmental Center for Telemedicine (CITEL) at the University of Bari. He serves as an Associate Editor for IEEE Transactions on Computational Social Systems, Soft Computing, and Computational Intelligence Magazine, and is a Guest Co-Editor of the Special Issue Explainability in Human-Centric AI in Information Systems Frontiers (Springer). He is a Senior Member of IEEE and a member of the IEEE Computational Intelligence Society Task Force on Explainable Fuzzy Systems. He regularly serves as reviewer, program committee member, and organizer for international conferences and journals in Artificial Intelligence and Computational Intelligence.

4.1. Program Committee

- Domenico Amato, Università degli Studi di Palermo, Italy
- Giustino Claudio Miglionico, University of Pisa, Italy
- Davide Cazzorla, University of Bari Aldo Moro, Italy
- Salvatore Calderaro, Università degli Studi di Palermo, Italy
- Daniel Leite, Aeronautics Institute of Technology (ITA), Brazil
- Fahad Ahmed Khokhar, University of Florence, Italy
- Raffaele Scaringi, University of Bari Aldo Moro, Italy
- Vincenzo Pasquadibisceglie, University of Bari Aldo Moro, Italy
- Luis Terán, Lucerne University of Applied Sciences and Arts, Switzerland
- Fabrizio Ruffini, University of Pisa, Italy
- Marcin Ostrowski, Systems Research Institute, Polish Academy of Sciences, Poland
- Ansh Kapadia, Systems Research Institute, Polish Academy of Sciences, Poland
- Abdelhakim Ziani, CentraleSupélec, France
- Filip Wichrowski, Systems Research Institute, Polish Academy of Sciences, Poland
- Zied Mnasri, University Tunis El Manar, Tunisia
- Muhammad Rashid, University of Turin, Italy
- Marek Reformat, University of Alberta, Canada
- Corrado Mencar, University of Bari Aldo Moro, Italy
- Alessandro Renda, University of Trieste, Trieste, Italy

- Souhir Ben Souissi, Bern University of Applied Sciences, Bern, Switzerland
- Pietro Ducange, University of Pisa, Pisa, Italy
- Katarzyna Miś, University of Silesia, Poland
- Natalia Díaz-Rodríguez, University of Granada, Spain
- Manuel Striani, University of Eastern Piedmont, Italy
- Alberto Fernández, University of Granada, Spain
- Patryk Żywica, Adam Mickiewicz University, Poland
- Gonzalo Nápoles, Tilburg University, The Netherlands
- Daniel Peralta, University of Granada, Spain
- Damiano Verda, Rulex Innovation Labs, Italy
- Sileshi Nibret Zeleke, Free University of Bozen-Bolzano, Italy
- Marco Cremaschi, University Milano-Bicocca, Milan, Italy

Acknowledgments

The EXPLIMED organizers would like to thank the organizing committee of the 35th International Joint Conference on Artificial Intelligence (IJCAI-ECAI 2026) for hosting the third edition of the workshop. We also extend our sincere gratitude to the organizers of the Workshop on Explainable AI in Space (EASi) for their valuable collaboration in organizing the Joint Workshop on Explainable AI in Space and Explainable Artificial Intelligence for the Medical Domain (EASi-EXPLIMED). This joint initiative provided an excellent opportunity to bring together researchers from different application domains, fostering interdisciplinary discussions, knowledge exchange, and networking, while encouraging new collaborations within the broader Explainable Artificial Intelligence community. Ga.C., Gi.C., and G.Z. are members of the CITEL - Centro Interdipartimentale della ricerca in Telemedicina, of the University of Bari Aldo Moro. Katarzyna Kaczmarek-Majer is supported from the project "ExplainMe: Explainable Artificial Intelligence for Monitoring Acoustic Features extracted from Speech" (FENG.02.02-IP.05-0302/23) carried out within the First Team programme of the Foundation for Polish Science co-financed by the European Union under the European Funds for Smart Economy 2021-2027 (FENG). AGV's research is supported by a Ph.D. fellowship funded under the Italian program 'D.M. n. 630, April 24, 2024' – NRRP, Mission 4, Component 2, Investment 3.3, co-funded by Bristol-Myers Squibb s.r.l. (CUP B91I24000160007).

Declaration on Generative AI

Generative AI tools, specifically OpenAI's ChatGPT and Grammarly, were used exclusively for grammar correction and language refinement. The authors conceived, wrote, and validated all content.

References

[1] C. Kim, S. U. Gadgil, S.-I. Lee, Transparency of medical artificial intelligence systems, Nature Reviews Bioengineering 4 (2026) 11–29.

[2] G. Casalino, G. Castellano, A. G. Valerio, G. Vessio, G. Zaza, Enhancing the explainability of neuro-fuzzy systems with large language models: A case study on eeg-based epileptic seizure classification, in: 2025 International Joint Conference on Neural Networks (IJCNN), IEEE, 2025, pp. 1–8.
[3] F. Ahmed, N. S. Naz, S. Khan, A. U. Rehman, W. M. Ismael, M. A. Khan, Explainable artificial intelligence (xai) in medical imaging: a systematic review of techniques, applications, and challenges, BMC Medical Imaging (2026).
[4] E. Tjoa, C. Guan, A survey on explainable artificial intelligence (xai): Toward medical xai, IEEE transactions on neural networks and learning systems 32 (2020) 4793–4813.
[5] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, et al., Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai, Information fusion 58 (2020) 82–115.
[6] N. Díaz-Rodríguez, J. Del Ser, M. Coeckelbergh, M. L. de Prado, E. Herrera-Viedma, F. Herrera, Connecting the dots in trustworthy artificial intelligence: From ai principles, ethics, and key requirements to responsible ai systems and regulation, Information Fusion (2023) 101896.
[7] L. Nannini, J. M. Alonso-Moral, A. Catalá, M. Lama, S. Barro, Operationalizing explainable artificial intelligence in the european union regulatory ecosystem, IEEE Intelligent Systems 39 (2024) 37–48.
[8] M. Mesinovic, T. Zhu, Clinical xai: A tutorial on concepts, methods, and modalities from linear models to llms, ACM Computing Surveys (2026).
[9] X. Kong, S. Liu, L. Zhu, Toward human-centered xai in practice: A survey, Machine Intelligence Research 21 (2024) 740–770.
[10] C. Panigutti, A. Beretta, D. Fadda, F. Giannotti, D. Pedreschi, A. Perotti, S. Rinzivillo, Co-design of human-centered, explainable ai for clinical decision support, ACM Transactions on Interactive Intelligent Systems 13 (2023) 1–35.
[11] S. Mohseni, N. Zarei, E. D. Ragan, A multidisciplinary survey and framework for design and evaluation of explainable ai systems, ACM Transactions on Interactive Intelligent Systems (TiiS) 11 (2021) 1–45.
[12] N. Hughes, Y. Jia, M. Sujan, T. Lawton, I. Habli, J. McDermid, Design principles for human-centred explainable ai: A scoping review, ACM Transactions on Interactive Intelligent Systems 16 (2026) 1–26.
[13] R. Alharith, J. Zhang, A. O. Ibrahim, Z. Wu, Concept-based explanation for deep vision models: A comprehensive survey on techniques, taxonomy, applications, and recent advances: R. alharith et al., International Journal of Computer Vision 134 (2026) 36.
[14] G. Casalino, G. Castellano, K. Kaczmarek-Majer, D. Schicchi, D. Taibi, G. Zaza, Evolving fuzzy classification for human-centered explainable learning analytics in virtual environments, Evolving Systems 16 (2025) 119.
[15] C. Michel-Delétie, M. K. Sarker, Neuro-symbolic methods for trustworthy ai: a systematic review, Neurosymbolic Artificial Intelligence (2024).
[16] E. Cappuccio, B. Kathirgamanathan, S. Rinzivillo, G. Andrienko, N. Andrienko, Integrating human knowledge for explainable ai, Machine Learning 114 (2025) 250.
[17] M. Nandan, S. Mitra, D. De, Graphxai: a survey of graph neural networks (gnns) for

explainable ai (xai), Neural Computing and Applications 37 (2025) 10949–11000.
[18] G. Zaza, R. Scaringi, G. Vessio, G. Castellano, Fuzzygnn: a graph neural network framework for enhanced fuzzy modeling, Neural Computing and Applications 38 (2026) 390.
[19] D. Mindlin, F. Beer, L. N. Sieger, S. Heindorf, E. Esposito, A.-C. N. Ngomo, P. Cimiano, Beyond one-shot explanations: a systematic literature review of dialogue-based xai approaches, Artificial Intelligence Review 58 (2025).
[20] O. Kaya, A. S. Shah, M. A. Karabulut, S. N. Karahan, M. S. Osmanca, N. Acır, Explainable artificial intelligence (xai): Concepts, applications, challenges, and future perspectives, IEEE Access (2026).
[21] J. Stefanowski, Challenges in evaluating explanation methods for static and evolving data, 2026. URL: https://arxiv.org/abs/2608.06351. arXiv:2608.06351.
[22] M. Mattioli, M. Pelillo, Scientific explanations in health sciences: Causality, trust, and epistemic adequacy, in: Proceedings of the Third Workshop on Explainable Artificial Intelligence for the Medical Domain, 17 August 2026, Bremen, Germany, co-located with the 35th International Joint Conference on Artificial Intelligence (IJCAI-ECAI 2026), volume — of *CEUR Workshop Proceedings*, 2026.
[23] K. Dobiczek, M. Mozolewski, S. Bobek, M. Szafarczyk, P. van Dam, G. J. Nalepa, Evaluating the role of data and explanation representations in clinical xai for conduction disorder classification, in: Proceedings of the Third Workshop on Explainable Artificial Intelligence for the Medical Domain, 17 August 2026, Bremen, Germany, co-located with the 35th International Joint Conference on Artificial Intelligence (IJCAI-ECAI 2026), volume — of *CEUR Workshop Proceedings*, 2026.
[24] L. Chen, T. E. M. Nordling, Explaining rhythmformer: A systematic xai analysis of periodic sparse attention for remote photoplethysmography, in: Proceedings of the Third Workshop on Explainable Artificial Intelligence for the Medical Domain, 17 August 2026, Bremen, Germany, co-located with the 35th International Joint Conference on Artificial Intelligence (IJCAI-ECAI 2026), volume — of *CEUR Workshop Proceedings*, 2026.
[25] Y. Wang, K. Kaczmarek-Majer, M. Chojnicki, P. Burchardt, Y. Ren, M. Z. Reformat, U-graphrag: Evidence-support estimation in a medical graphrag system, in: Proceedings of the Third Workshop on Explainable Artificial Intelligence for the Medical Domain, 17 August 2026, Bremen, Germany, co-located with the 35th International Joint Conference on Artificial Intelligence (IJCAI-ECAI 2026), volume — of *CEUR Workshop Proceedings*, 2026.
[26] A. Tarek, H. Saber, H. Elghonemy, Y. Afify, T. Basha, O. S. Bhatti, A. M. Selim, H. M. T. Alam, D. Sonntag, Trace: A concept bottleneck model for longitudinal 3d glioblastoma response assessment, in: Proceedings of the Third Workshop on Explainable Artificial Intelligence for the Medical Domain, 17 August 2026, Bremen, Germany, co-located with the 35th International Joint Conference on Artificial Intelligence (IJCAI-ECAI 2026), volume — of *CEUR Workshop Proceedings*, 2026.
[27] E. Barbu, R. Krjutškov, A. Uusküla, Evaluating conversational explainability in healthcare: Comparing the explainability assistant and a traditional xai dashboard, in: Proceedings of the Third Workshop on Explainable Artificial Intelligence for the Medical Domain, 17 August 2026, Bremen, Germany, co-located with the 35th International Joint Conference on Artificial Intelligence (IJCAI-ECAI 2026), volume — of *CEUR Workshop Proceedings*,

2026.

[28] J. Rymarski, A. Rempała, B. Sobieski, P. Biecek, Explaining biomedclip with weighted banzhaf interactions supported by tree-gram parsing, in: Proceedings of the Third Workshop on Explainable Artificial Intelligence for the Medical Domain, 17 August 2026, Bremen, Germany, co-located with the 35th International Joint Conference on Artificial Intelligence (IJCAI-ECAI 2026), volume — of *CEUR Workshop Proceedings*, 2026.

[29] B. İlgen, A. Shahid, D. Kühnert, Interpretable flavivirus classification under low-coverage conditions via erf-aware k-mer attribution, in: Proceedings of the Third Workshop on Explainable Artificial Intelligence for the Medical Domain, 17 August 2026, Bremen, Germany, co-located with the 35th International Joint Conference on Artificial Intelligence (IJCAI-ECAI 2026), volume — of *CEUR Workshop Proceedings*, 2026.

[30] D. Babuc, Histodyn-biogater: Biologically interpretable dynamic gating architecture for computational pathology, in: Proceedings of the Third Workshop on Explainable Artificial Intelligence for the Medical Domain, 17 August 2026, Bremen, Germany, co-located with the 35th International Joint Conference on Artificial Intelligence (IJCAI-ECAI 2026), volume — of *CEUR Workshop Proceedings*, 2026.

[31] C. Listone, S. Romano, N. Spanoudakis, A. Murano, Argo: An argumentative framework for guided medical outcomes, in: Proceedings of the Third Workshop on Explainable Artificial Intelligence for the Medical Domain, 17 August 2026, Bremen, Germany, co-located with the 35th International Joint Conference on Artificial Intelligence (IJCAI-ECAI 2026), volume — of *CEUR Workshop Proceedings*, 2026.