A Human-Centered Explainability Framework for Latent Clustering Based on NMF, Fuzzy Logic, and LLMs

Gabriella Casalino*, Giovanna Castellano*, Katarzyna Kaczmarek-Majer†, Giovanni Narracci*, Vittorio Nitti*, Francesco Pagliaro*, Alberto G. Valerio*, Gianluca Zaza*

* Computer Science Dept., University of Bari Aldo Moro – Bari, Italy.
{*gabriella.casalino, giovanna.castellano, gianluca.zaza*}*@uniba.it*
†Systems Research Institute Polish Academy of Sciences – Warsaw, Poland.
k.kaczmarek@ibspan.waw.pl

***Abstract*—Explaining unsupervised models in a human-centered manner remains challenging, particularly when relying on latent representations. While Non-negative Matrix Factorization (NMF) provides structured latent factors, their semantic interpretation is often limited. Conversely, Large Language Models (LLMs) can generate fluent explanations but require semantically grounded inputs to ensure faithfulness.**

In this work, we propose a modular framework integrating NMF, fuzzy semantic modeling, and LLM-based explanation generation to support human-centered explanations of clustering results. Fuzzy logic maps latent factors and cluster representatives into linguistically meaningful concepts, which guide LLMs through structured prompting. We evaluate six prompting strategies with increasing semantic enrichment across eleven open-source LLMs.

The framework is validated on an acoustic analysis task related to bipolar disorder, where interpretability is critical for expert understanding. Results show that combining concept-based representations with structured prompting improves the accessibility and reliability of LLM-generated explanations in domain-sensitive contexts.

***Index Terms*—Non-Negative Matrix Factorization, Fuzzy Logic, Human-Centered Explainable Artificial Intelligence, Large Language Models, Bipolar Disorder, Mental Health Disease**

I. Introduction

The increasing adoption of complex, data-driven models in critical application domains has intensified the demand for effective Explainable Artificial Intelligence (XAI) solutions [1]–[3]. However, even when such models provide interpretable or partially interpretable representations, the resulting explanations are often not aligned with users' needs, background, or cognitive expectations, limiting their usefulness in real-world decision-making scenarios [4]. To address this issue, human-centered XAI has emerged as a paradigm that emphasizes explanations that are not only technically faithful but also cognitively accessible and meaningful to end users [5].

Most existing XAI approaches focus on supervised learning, in which explanations are typically expressed as feature relevance or decision boundaries. In contrast, explainability in unsupervised models remains comparatively underexplored [7]. The absence of target labels makes it difficult to define meaningful reference points, posing challenges both for generating and evaluating explanations that are understandable to humans.

In this context, concept-based approaches are a promising direction, as they rely on higher-level, human-meaningful, and coherent concepts shared across the dataset [8], [9] to bridge the gap between low-level numerical representations and human interpretation.

In our previous work [6], we introduced a concept-based representation framework that combines Non-negative Matrix Factorization (NMF) and fuzzy logic to map numerical latent factors to linguistically meaningful concepts. While this approach improved interpretability, the resulting descriptions remained limited in expressiveness and lacked the contextual richness required for fully human-centered explanations.

In this work, we further extend our framework by integrating Large Language Models (LLMs) to enhance the expressiveness of concept-based representations. Motivated by their ability to generate fluent and context-aware natural-language explanations [10], [11], we leverage LLM-based generation to transform structured representations into fully articulated explanations. To this end, we propose a novel explanation pipeline that combines NMF-based representations, fuzzy semantic modeling, and LLM-based generation to produce human-centered explanations of clustering results while preserving fidelity to the underlying data.

A key contribution of this work lies in the systematic comparison and evaluation of LLMs within this pipeline. Given the variability of LLM outputs with respect to both model architecture and prompt design, assessing their impact on explanation quality is essential to ensure reliability and consistency in human-centered XAI. To this end, we conduct a comprehensive experimental study involving six prompting strategies and eleven open-source LLMs, analyzing how different configurations affect multiple dimensions of explanation quality. This evaluation provides methodological insights into how to effectively select and use LLMs for explanation generation, addressing a gap in current XAI research where such

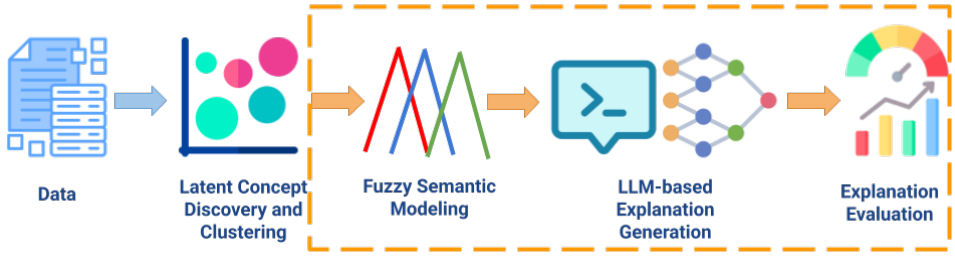

Fig. 1: Overview of the proposed explanation pipeline. The orange-highlighted component denotes the module newly introduced with respect to [6].

comparative analyses are still limited.

The proposed framework is evaluated in the context of acoustic pattern analysis for bipolar disorder, a domain in which explainability is crucial to support expert interpretation and foster trust in data-driven clinical insights.

The remainder of the paper is organized as follows. Section II presents the proposed workflow. Section III describes the experimental setup, while Section IV reports and discusses the obtained results. Finally, Section V concludes the paper and outlines future research directions.

II. Proposed Workflow

We propose a modular and interpretable pipeline, building upon [6] and further extending it with additional modules for explanation and evaluation, to produce high-quality, human-centered explanations from complex data representations using LLMs.

The overall process comprises four main modules, each addressing a specific stage of the explanation generation workflow, as illustrated in Fig. 1, where the newly introduced components are highlighted. First, the input data are clustered, and latent interpretable concepts are automatically extracted using NMF. Second, fuzzy semantic modeling translates these representations into linguistic descriptions that reflect human-like perception. Third, LLMs are employed to generate coherent natural-language explanations grounded in the structured semantic information derived from the previous modules. Finally, a dedicated evaluation module assesses the quality of the generated explanations along multiple complementary dimensions, including linguistic quality, semantic coherence, information coverage, and efficiency.

A. Latent Concept Discovery and Clustering

In this work, Non-negative Matrix Factorization (NMF) [12] is adopted to extract latent, interpretable concepts from the input data and to support clustering in a reduced representation space. NMF decomposes the non-negative data matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$ into two non-negative matrices, $\mathbf{W} \in \mathbb{R}^{m \times k}$ and $\mathbf{H} \in \mathbb{R}^{k \times n}$, such that $\mathbf{X} \approx \mathbf{WH}$, where $k \ll \min(m, n)$:

$$\min_{\mathbf{W},\mathbf{H}} \|\mathbf{X} - \mathbf{WH}\|_F^2 \quad \text{s.t.} \quad \mathbf{X} \geq \mathbf{0},\ \mathbf{W} \geq \mathbf{0},\ \mathbf{H} \geq \mathbf{0}. \tag{1}$$

Within the proposed pipeline, the matrix $\mathbf{W}$ encodes the latent factors in terms of the original features, enabling their direct interpretation as high-level concepts, while $\mathbf{H}$ provides a reduced representation of data instances in the latent space. This representation is exploited to group similar instances by assigning each sample to the latent factor that most strongly contributes to its reconstruction.

As a result, NMF provides both a concept-based description of the data and a clustering structure grounded in these concepts, which serve as the basis for the subsequent semantic modeling and explanation generation steps.

B. Fuzzy Semantic Modeling

In the second module, fuzzy logic is employed to transform the numerical representations derived from NMF into linguistically interpretable descriptions, thus enhancing their suitability for human-centered explanations. Building on the formalism introduced in [6], both latent factors (encoded in $\mathbf{W}$) and cluster-level representations (derived from $\mathbf{H}$) are mapped into fuzzy linguistic variables.

Let $\mathcal{F} = \{f_1, f_2, \ldots, f_m\}$ denote the set of original features. For each feature f_i, a fuzzy linguistic variable is defined together with an associated fuzzy partition $\{A^{(is)} \mid s = 1, \ldots, g\}$, where g denotes the granularity. Each coefficient w_{ij} is then mapped to the linguistic term that maximizes its membership degree:

$$V^{(i)} = \arg\max_{s} \{\mu_{A^{(is)}}(w_{ij})\} \tag{2}$$

where $\mu_{A^{(is)}}(w_{ij})$ denotes the membership degree of w_{ij} to the fuzzy set $A^{(is)}$.

This mapping enables the generation of linguistic descriptions of latent factors in terms of the original features, providing a semantic characterization of the extracted concepts as:

$$\text{The influence on } \mathbf{w}_{*j} \text{ is :}$$
$$V^{(1)} \text{ for } \mathbf{f}_1,\ V^{(2)} \text{ for } \mathbf{f}_{2*},\ \ldots,\ V^{(m)} \text{ for } \mathbf{f}_m.$$

A similar fuzzification process is applied to cluster representatives computed in the latent space. Specifically, for each cluster C_j, a prototype $\tilde{\mathbf{h}}_{*j}$ is derived and mapped into fuzzy linguistic terms by assigning each component to the fuzzy set with the highest membership degree:

$$V^{(i)} = \arg\max_{s'} \left\{ \mu_{A^{(is')}}(\tilde{h}_{ij}) \right\}. \quad (3)$$

This yields a human-readable linguistic characterization of each cluster in terms of the contributions of its latent factors.

C. LLM-based explanation generation

Building on the structured and semantically enriched representations obtained from the previous modules, the third component of the pipeline focuses on generating coherent and human-centered natural-language explanations.

Specifically, LLMs are employed to synthesize explanations that integrate both the semantic characterization of latent factors and the linguistic description of clusters. Unlike standard LLM-based approaches, which often rely on raw or weakly structured inputs, the proposed method frames the generation process as a guided and structured explanation task grounded in concept-based and fuzzy linguistic representations. This design ensures that the generated explanations remain faithful to the underlying data while preserving fluency and expressiveness.

To this end, prompt engineering strategies are adopted to guide the LLM in combining cluster-level and factor-level information into coherent explanations. The prompts explicitly encode the outputs of the previous modules, enabling controlled generation and reducing the risk of ungrounded or inconsistent content.

In this work, LLMs are used in a zero-shot setting to assess their ability to generate accurate, coherent, and semantically consistent explanations from the provided structured information, without task-specific fine-tuning. This design highlights the potential of LLMs as components of explainable pipelines when properly guided by interpretable representations.

D. Explanation Evaluation

The final module of the proposed pipeline focuses on the quantitative assessment of the generated explanations. In this work, explanation evaluation is treated as an integral component of the framework, enabling a systematic analysis of multiple dimensions that characterize the quality of LLM-generated explanations.

Specifically, the generated explanations are evaluated along complementary criteria, including lexical diversity, readability, coherence, information coverage, and generation efficiency. *Lexical diversity* quantifies the variability of the vocabulary used in explanations, whereas *readability* assesses their linguistic accessibility. *Coherence* evaluates the semantic consistency between consecutive sentences, and *information coverage* measures the extent to which the generated explanations reflect the relevant information provided in the input. Finally, generation efficiency accounts for the *time* required to produce the explanations.

The evaluation follows the protocol introduced in [13], to which the reader is referred for full details on the adopted metrics.

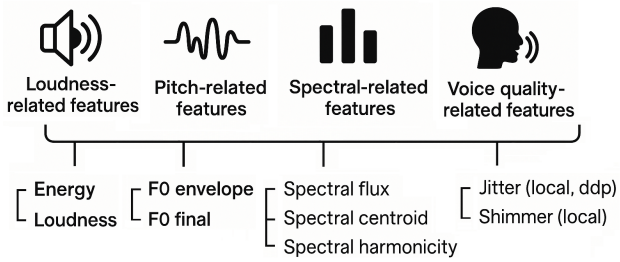

Fig. 2: Hierarchical organization of original features into derived meta-features [14].

III. Experimental settings

A. Data description

The dataset consists of daily-aggregated acoustic representations obtained by summarizing patients' speech activity over time. From the set of descriptors provided by the OpenSMILE library [15], we selected ten low-level acoustic features (Fig. 2) capturing complementary aspects of vocal behavior relevant to bipolar disorder mood fluctuations, while limiting dimensionality to favor interpretability [16].

These features are used as input to NMF to extract latent patterns and identify groups of patients with similar acoustic profiles. To further enhance interpretability and facilitate understanding by medical experts, we adopt an abstraction strategy introduced in [14], grouping low-level features into four high-level semantic categories (Fig. 2).

These high-level features are used to inject domain-relevant semantic information into the LLMs, supporting the generation of explanations that can be more easily related to patients' affective and behavioral states.

B. Configuration of NMF and Fuzzy Components

We analyzed the impact of different NMF initialization strategies and optimization algorithms, given the known sensitivity of NMF to these choices. Specifically, we compared random and SVD-based initializations, together with Coordinate Descent, Multiplicative Update, and Alternating Least Squares optimization methods [12]. The optimal configuration and latent dimensionality ($k \in [2, 10]$) were selected based on reconstruction error and Silhouette index. The best results were obtained using SVD-based initialization with the Multiplicative Update algorithm and $k = 2$.

To describe both original features and latent factors, we adopted three Gaussian fuzzy sets uniformly distributed in the $[0, 1]$ interval. Gaussian membership functions were selected to ensure smooth transitions between linguistic terms, reflecting gradual variations in acoustic patterns.

We also evaluated alternative configurations by varying the number of fuzzy sets (from 3 to 5) and their shapes (triangular vs. Gaussian), observing no significant differences in the generated explanations. Therefore, we selected three Gaussian fuzzy sets as the most compact and interpretable configuration.

C. Large Language Models

A comprehensive experimental study was conducted involving 11 LLMs, selected to cover a diverse range of model families, sizes, and architectural paradigms, while ensuring compatibility with local deployment. The analysis aims to investigate both intra-family variability (models within the same family with different configurations) and inter-family differences across heterogeneous LLM frameworks.

All models are open-source and suitable for on-premise execution. The evaluated models include representatives from the LLaMA family (LLaMA 3 and LLaMA 4 variants) [17]–[19], as well as models from other recent architectures, including Allam-2-7B [20], DeepSeek R1 Distill LLaMA-70B [21], Qwen-QWQ-32B [22], and inference-optimized Compound models [23], [24].

D. Prompt Engineering

To guide the LLMs in the explanation task, we adopt a zero-shot prompt engineering strategy, leveraging the intrinsic knowledge encoded in the models. The prompts are designed to combine domain knowledge with the structured representations derived from NMF and fuzzy modeling, enabling the generation of explanations that are both accessible and informative.

We evaluate six prompting strategies sharing a common informational backbone. In all configurations, the LLM receives: (i) the fuzzy linguistic characterization of clusters in terms of latent factors, and (ii) the semantic description of latent factors with respect to the original feature space. Based on this input, the model is instructed to generate concise natural-language explanations linking latent factors, acoustic features, and clinical interpretations.

The evaluated prompting strategies are:

P1 - Zero-Shot Prompting: baseline generation directly from fuzzy descriptions;

P2 - Persona Pattern: the LLM adopts the role of a senior psychiatrist to tailor explanations to the target audience, enhancing relevance and communicative effectiveness in line with Human–Computer Interaction principles;

P3 - Persona + Fact-Checking: extends P2 by encouraging fact-based and grounded explanations, to promote transparency and trustworthiness;

P4 - Zero-Shot + Context: enriches P1 with contextual information on low-level acoustic features;

P5 - Persona + Context: combines persona-based prompting with contextual enrichment;

P6 - Persona + Context + Meta Feature: further integrates high-level meta-features to support more abstract and clinically meaningful explanations.

IV. Results

A. Semantic Interpretation of Latent Factors and Clusters

The explanations of the latent factors obtained through NMF and fuzzy semantic modeling are as follows.

> The influence on LF1 is: **High** for **`f0_envelope`**, **High** for **`f0_final`**, Low for `jitter_ddp`, Low for `jitter_local`, Low for `loudness`, **High** for **`spectral_centroid`**, Low for `spectral_flux`, Low for `spectral_harmonicity`, Low for `energy`, Low for `shimmer_local`.

The semantic meaning of LF1 can be inferred from the acoustic features that are predominantly represented (in bold). LF1 reflects prosodic dynamics and vocal brightness, with strong intonational variability and increased pitch and spectral energy, which are commonly associated with emotional activation.

> The influence on LF2 is: Low for `f0_envelope`, Low for `f0_final`, **High** for **`jitter_ddp`**, **High** for **`jitter_local`**, **High** for **`loudness`**, Low for `spectral_centroid`, **High** for **`spectral_flux`**, **High** for **`spectral_harmonicity`**, **High** for **`energy`**, **High** for **`shimmer_local`**.

In contrast, LF2 is primarily characterized by increased micro-prosodic variability, reflected by high jitter and shimmer values, together with elevated loudness, energy, and spectral flux. This combination indicates an energetically intense yet unstable vocal profile, associated with reduced control over pitch and amplitude and increased spectral irregularity. Such acoustic patterns are commonly linked to vocal strain, emotional distress, and impaired motor control of speech.

As with the latent factors, cluster-level explanations were derived from the previously described latent factors. Although these explanations are informative and interpretable, their communicative effectiveness can be further enhanced. Therefore, they are used as input to LLMs in the subsequent steps to improve clarity and accessibility for end users. The following example illustrates the explanation of cluster 0.

> Samples in cluster 0 are characterized by: **High** values of **LF1** and **Low** values of **LF2**.

B. Global Analysis Across Prompting Strategies

Table I reports the average performance (mean and standard deviation) across all evaluated LLMs and clusters for each prompting strategy, considering normalized scores for diversity, readability, coherence, and coverage, together with the average generation time. Overall, the results reveal a clear trend whereby increasing prompt complexity leads to more diverse and informative explanations, at the cost of higher computational time.

The zero-shot baseline (P1) achieves competitive coherence and coverage with the lowest generation time. Persona-based prompting (P2) and its fact-checking extension (P3) primarily improve diversity while maintaining high coherence and coverage. The inclusion of contextual information (P4 and P5) yields more substantial gains in diversity and coverage, confirming the importance of grounding explanations in feature-level semantics. Finally, the most enriched strategy (P6) attains the highest diversity and coverage scores, but also exhibits the highest generation time.

Readability values are relatively low but remain stable across all prompting strategies. This suggests that the generated explanations consistently preserve a technically dense and information-rich structure, which is expected in a domain-specific setting. The limited variability indicates that readabil-

TABLE I: Average performance across models and clusters for each prompt pattern.

Prompt	Diversity	Readability	Coherence	Coverage	Time
P1	0.47 ± 0.25	0.29 ± 0.20	0.58 ± 0.31	0.57 ± 0.27	5.40 ± 2.69
P2	0.54 ± 0.29	0.29 ± 0.23	0.58 ± 0.31	0.43 ± 0.28	6.20 ± 3.22
P3	0.58 ± 0.28	0.26 ± 0.23	0.57 ± 0.32	0.43 ± 0.29	6.56 ± 3.35
P4	0.61 ± 0.29	0.25 ± 0.22	0.59 ± 0.31	0.53 ± 0.25	7.17 ± 3.59
P5	0.64 ± 0.30	0.24 ± 0.23	0.53 ± 0.29	0.54 ± 0.24	7.10 ± 3.55
P6	0.67 ± 0.29	0.23 ± 0.22	0.55 ± 0.28	0.58 ± 0.25	7.47 ± 3.79

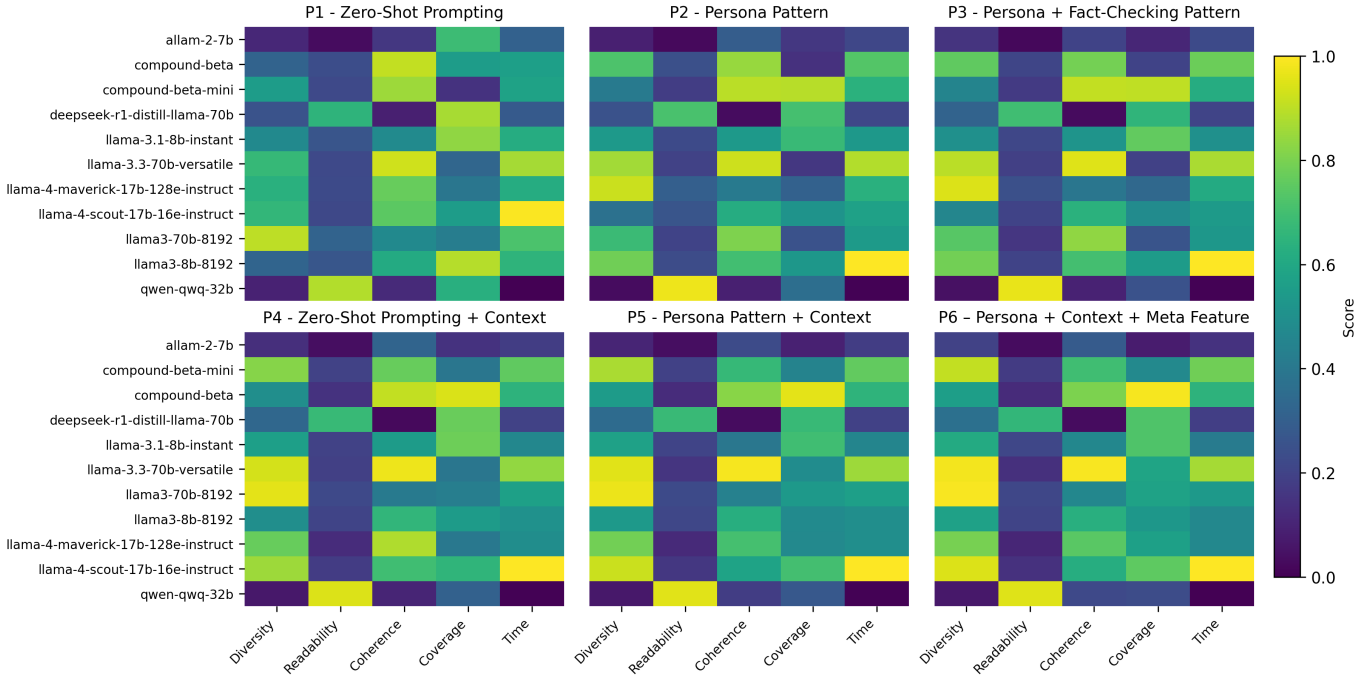

Fig. 3: Comparison of explanation quality across prompting strategies.

ity is mainly driven by the underlying structured representations rather than by prompt formulation.

Notably, relatively high standard deviations are observed across most metrics and prompting strategies. This variability can be attributed to three main factors. First, the aggregation across multiple LLMs with heterogeneous architectures, sizes, and training data introduces substantial variability in generation behavior. Second, different clusters are characterized by varying levels of semantic complexity and feature distributions, which affect the difficulty of generating coherent and comprehensive explanations. Third, prompt enrichment does not impact all models uniformly, leading to divergent responses depending on the model–prompt interaction.

C. *Model-Level Trends*

Figure 3 provides a detailed model-wise analysis of the prompting strategies, highlighting how different LLMs respond to increasing prompt complexity across the evaluated metrics. Overall, no single LLM consistently dominates across all prompts and measures; rather, performance is highly prompt-dependent. Larger and more recent models (e.g., LLaMA-3/4 and Compound-based models) tend to achieve higher diversity, particularly under enriched prompts (P4–P6), while smaller or distilled models show more stable behavior under simpler configurations (P1–P2). Coherence appears to be the most robust metric, remaining relatively high across most models and prompts, whereas readability exhibits limited variation and is less sensitive to prompt enrichment. In contrast, diversity and coverage benefit the most from contextual and meta-feature augmentation, but only for a subset of LLMs, indicating that these metrics are not uniformly improved across models. Regarding efficiency, the normalized and inverted time scores reveal that simpler prompts favor faster generation across all LLMs, while structured prompts (especially P6) introduce a clear trade-off between richer explanations and reduced efficiency. Overall, the figure suggests that different LLMs excel on different metrics and under different prompting strategies, reinforcing the need for joint prompt–model selection rather than assuming a universally optimal LLM.

V. Conclusion

In this work, we proposed a modular human-centered explainability framework that integrates NMF-based latent clustering, fuzzy semantic modeling, and LLM-based explanation generation. The results show that fuzzy concept-based representations provide a semantically grounded and interpretable intermediate layer, enabling LLMs to generate coherent and meaningful explanations of unsupervised clustering outcomes.

The experimental analysis highlights that explanation quality strongly depends on both the prompting strategy and the selected LLM, with enriched prompts improving diversity and coverage at the cost of higher computational effort. Overall, the findings confirm the importance of combining structured semantic representations with carefully designed prompts to support accessible and reliable explanations in domain-sensitive applications such as bipolar disorder analysis. Future developments will move toward a fully human-centered XAI perspective by actively involving end users in evaluating the generated explanations, both to assess their usefulness and to co-design explanation processes that better reflect users' needs, expectations, and domain knowledge. In addition, given the use of LLMs for explanation generation, future work will explicitly address the detection and mitigation of hallucinations, in order to ensure the faithfulness and reliability of the generated explanations, especially in sensitive application domains.

ACKNOWLEDGMENT

During the preparation of this work, the authors used GPT-5.3 (ChatGPT, OpenAI) and Grammarly exclusively for language refinement (grammar and writing style). No AI tools were used to generate scientific content, ideas, or interpretations. All content was developed, reviewed, and approved by the authors, who take full responsibility for the manuscript.

This paper has been supported by the "INdAM - GNCS Project" CUP_E53C24001950001 . Ga.C, G.C, A.V., and G.Z. are INdAM GNCS research group members. A Ph.D. fellowship funds A.G.V.'s research within the Italian "D.M. n. 630, April 24, 2024" – under the NRRP, Mission 4, Component 2, Investment 3.3 – Ph.D. project "Explainability of Artificial Intelligence systems for applications in the medical field", co-supported by Bristol-Myers Squibb s.r.l. (CUP B91I24000160007). Katarzyna Kaczmarek-Majer is supported by the project "ExplainMe: Explainable Artificial Intelligence for Monitoring Acoustic Features extracted from Speech" (FENG.02.02-IP.05-0302/23), which is 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).

REFERENCES

[1] L. Nannini, J. M. Alonso-Moral, A. Catalá, M. Lama, and S. Barro, "Operationalizing explainable artificial intelligence in the european union regulatory ecosystem," IEEE Intelligent Systems, vol. 39, no. 4, pp. 37–48, 2024.

[2] N. Díaz-Rodríguez, J. Del Ser, M. Coeckelbergh, M. L. de Prado, E. Herrera-Viedma, and F. Herrera, "Connecting the dots in trustworthy artificial intelligence: From ai principles, ethics, and key requirements to responsible ai systems and regulation," Information Fusion, p. 101896, 2023.

[3] L. Longo, M. Brcic, F. Cabitza, J. Choi, R. Confalonieri, J. Del Ser, R. Guidotti, Y. Hayashi, F. Herrera, A. Holzinger et al., "Explainable artificial intelligence (xai) 2.0: A manifesto of open challenges and interdisciplinary research directions," Information Fusion, vol. 106, p. 102301, 2024.

[4] W. Yang, Y. Wei, H. Wei, Y. Chen, G. Huang, X. Li, R. Li, N. Yao, X. Wang, X. Gu et al., "Survey on explainable ai: From approaches, limitations and applications aspects," Human-Centric Intelligent Systems, vol. 3, no. 3, pp. 161–188, 2023.

[5] Y. Rong, T. Leemann, T.-T. Nguyen, L. Fiedler, P. Qian, V. Unhelkar, T. Seidel, G. Kasneci, and E. Kasneci, "Towards human-centered explainable ai: A survey of user studies for model explanations," IEEE transactions on pattern analysis and machine intelligence, vol. 46, no. 4, pp. 2104–2122, 2023.

[6] G. Casalino, G. Castellano, K. Kaczmarek-Majer, and G. Zaza, "Enhancing non-negative matrix factorization interpretability with fuzzy representations," in 2025 IEEE International Conference on Fuzzy Systems (FUZZ). IEEE, 2025, pp. 1–6.

[7] J. Kauffmann, M. Esders, L. Ruff, G. Montavon, W. Samek, and K.-R. Müller, "From clustering to cluster explanations via neural networks," IEEE transactions on neural networks and learning systems, vol. 35, no. 2, pp. 1926–1940, 2022.

[8] A. Ghorbani, J. Wexler, J. Y. Zou, and B. Kim, "Towards automatic concept-based explanations," Advances in neural information processing systems, vol. 32, 2019.

[9] E. Poeta, G. Ciravegna, E. Pastor, T. Cerquitelli, and E. Baralis, "Concept-based explainable artificial intelligence: A survey," ACM Computing Surveys, 2023.

[10] E. Paraschou, I. Arapakis, S. Yfantidou, S. Macaluso, and A. Vakali, "Mind the xai gap: A human-centered llm framework for democratizing explainable ai," in World Conference on Explainable Artificial Intelligence. Springer, 2025, pp. 113–137.

[11] A. G. Valerio, K. Trufanova, S. De Benedictis, G. Vessio, and G. Castellano, "From segmentation to explanation: Generating textual reports from mri with llms," Computer Methods and Programs in Biomedicine, vol. 270, p. 108922, 2025.

[12] N. Gillis, Nonnegative matrix factorization. SIAM, 2020.

[13] G. Casalino, G. Castellano, D. Margherita, A. G. Valerio, G. Vessio, and G. Zaza, "Exploring the expressive power of large language models in neuro-fuzzy system explainability: A study on eeg-based seizure detection," in Proceedings of the Second Workshop on Explainable Artificial Intelligence for the medical domain - 25-30 October 2025, Bologna, Italy, co-located with 28th European Conference on Artificial Intelligence - ECAI 2025, ser. CEUR Workshop Proceedings, vol. —, 2025.

[14] K. Kaczmarek-Majer, G. Casalino, G. Castellano, O. Hryniewicz, and M. Dominiak, "Explaining smartphone-based acoustic data in bipolar disorder: Semi-supervised fuzzy clustering and relative linguistic summaries," Information Sciences, vol. 588, pp. 174–195, 2022.

[15] F. Eyben, F. Weninger, F. Gross, and B. Schuller, "Recent developments in opensmile, the munich open-source multimedia feature extractor," in Proc. of the 21st ACM Int. Conf. on Multimedia, 2013, pp. 835–838.

[16] K. Kaczmarek-Majer, M. Dominiak, A. Z. Antosik, O. Hryniewicz, O. Kamińska, K. Opara, J. Owsiński, W. Radziszewska, M. Sochacka, and L. Swiecicki, "Acoustic features from speech as markers of depressive and manic symptoms in bipolar disorder: A prospective study," Acta Psychiatrica Scandinavica, 2024.

[17] G. Aaron and et al., "The llama 3 herd of models," 2024.

[18] Meta, "Llama 4 Maverick 17B-128E-Instruct-FP8," https://www.llama.com/docs/llama-4-maverick, 2025, version: 17B active parameters, 400B total parameters, 128 experts.

[19] ——, "Llama 4: A New Foundation Model," https://www.llama.com/docs/llama-4-scout, 2025, version: Scout (17B).

[20] B. M Saiful and et al., "ALLam: Large language models for arabic and english," in The Thirteenth International Conference on Learning Representations, 2025.

[21] L. Aixin and et al., "Deepseek-v3 technical report," 2025.

[22] Y. An and et al., "Qwen3 technical report," 2025.

[23] Groq, "Compound Beta, a Compound AI System," https://console.groq.com/docs/compound/systems/compound-beta, 2025, version: 2025-07-23 (Stable); Powered by Llama 4 Scout and Llama 3.3 70B.

[24] ——, "Compound Beta Mini, a Compound AI System," https://console.groq.com/docs/compound/systems/compound-beta-mini, 2025, version: 2025-07-23 (Stable); Powered by Llama 4 Scout and Llama 3.3 70B.