National Journal of Speech and Audio Processing, ISSN: 3107-5258 Vol. 1, No. 3, 2025 (pp. 46-53)
RESEARCH ARTICLE
ECEJOURNALS.IN

Explainable Artificial Intelligence in Speech and Audio Processing: Enhancing Interpretability, Fostering Trust, and Addressing Deployment Challenges

Daniel Lalit[1]*, Jinfe Regash[2]

[1]*Department of ECE and CpE, Ateneo de Naga University, Naga City, Bicol Region, Philippines*
[2]*Electrical and Computer Engineering Addis Ababa University Addis Ababa, Ethiopia*

KEYWORDS:
Explainable AI,
Speech Recognition,
Audio Processing,
Interpretability,
Trust,
Model Explainability,
Deployment Challenges.

ARTICLE HISTORY:
Submitted : 16.04.2025
Revised : 07.05.2025
Accepted : 23.07.2025

https://doi.org/10.17051/NJSAP/01.03.06

Abstract

The speed at which deep learning has found an application in speech and audio processing has demonstrated significant performance improvements across several applications such as the automatic speech recognition (ASR), speaker verification, emotion recognition, and audio event detection. Nevertheless, the black boxed, opaque nature of state of the art training poses some serious problems of transparency, interpretability, and trust of users, especially in safety critical and privacy sensitive areas. Explainable Artificial Intelligence (XAI) is a way forward in mitigating such problems as it offers inroads into the inner workings of AI systems. This paper works as a review of the existing XAI techniques applied to speech and audio processing and classifies the methods into model-specific and model-agnostic, explaining the most acceptable metrics of their interpretability. We investigate how explainability can increase trustworthiness, promote regulatory compliance, assist with debugging the systems and reduce bias. The issues of deployment, including real-time interpretability in the face of computational constraints, cross-lingual robustness, and human- machine communication, are critically assessed. We also find gaps in the research which can be explored further such as low-latency explanation generation, multimodal explainability and privacy-preserving explanation mechanisms. Lastly, we lay out a roadmap to the inclusion of XAI in next-generation speech and audio systems and ways to promote the deployment of responsible, transparent, and trusted AI in both business and mission-critical scenarios. Three case studies show that XAI is effective in detecting bias, confirming feature explanations and resolving errors in ASR, emotion recognition, and audio event classification.

Author's e-mail: daniel.lma@gmail.com, jin.regash@aait.edu.et

How to cite this article: Lalit D, Regash J. Explainable Artificial Intelligence in Speech and Audio Processing: Enhancing Interpretability, Fostering Trust, and Addressing Deployment Challenges. National Journal of Speech and Audio Processing, Vol. 1, No. 3, 2025 (pp. 46-53).

Introduction

Advances in deep learning during the past eight years have resulted in a revolution in the way speech and audio are processed by computers, with recent developments achieving state-of-the-art accuracy in automatic speech recognition (ASR),[1] speaker verification,[2] emotion detection[3] and acoustic scene labeling.[4] Neural networks, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), use of Transformer architecture, and self-supervised networks, including wav2vec 2.0, have proven accurate across multiple datasets, and across diverse acoustic environments.[5] Nevertheless, these developments have caused major issues with regards to transparency, interpretability, as well as trustworthiness due to the black-jacket of this kind of models. Model prediction can lead to life-changing consequences in systems operating in high-stakes areas like healthcare, finance, law enforcement, and assistive technologies, thus explaining a less critical but inevitably needed feature than newly established best practices.[6, 7] Legal requirements, such as the General Data Protection Regulation (GDPR) right to explanation also heighten the need to have transparent AI systems. Explainable Artificial Intelligence (XAI) has recently increased the interest in the computer vision and natural language processing domains, whereas its

implementation in the speech and audio context has been relatively under explored. Current literature and studies tend to pursue performance optimisation, but not much attention has been paid to model interpretability, bias monitoring, and trust modeling as perceived by the user.[8, 9] Furthermore, the existing XAI techniques on audio tend to be impractical in the real-time setting, have problems with multilingual scalability, and offer explainability that cannot be comprehended by non-experts.

In this paper, the authors will fill those gaps by exploring how XAI can effectively be added to systems processing speech and audio in a systematic manner. We look at three pillars of foundations:

1. Increasing Interpretability -improve interpretation of model internals and outputs,
2. Building Trust - creating greater belief in system predictions on the part of the user, and
3. Deployment Challenges Deployment Solutions: This is addressing the issue of real-time, scalable and ethically acceptable integration of solutions.

The rest of the paper is structured as follows: In Section 2, we review background material and related work in the application of XAI to audio; Section 3 discusses interpretability methods; Section 4 speaks to trust assistance mechanisms; Section 5 discusses implementation challenges; Section 6 outlines open research opportunities; and Section 7 provides a closing integration strategy proposal.

Background and Related Work

Speech and Audio Processing Applications

With deep learning, speech and audio processing technologies have taken levels where high-performance solutions can be implemented to solve many real-life situations:

- Automatic Speech Recognition (ASR): Tasks wherein perfectly transcribed speech is converted to text with a high degree of accuracy, even in relatively noisy settings, include systems like Google Speech-to-Text, Mozilla DeepSpeech, and self-supervised models like wav2vec 2.0.[10, 11]
- Speaker Verification and Identification: Deep embedding feature x-vectors, d-vectors, are extensively applied to authenticate biometrics, forensic and personalized voice services.[12]
- Emotion and Sentiment Recognition: Prosody- and spectrum-based prosodies recognized by CNNs or attention-based networks allow detecting emotions to monitor healthcare, automate customer services and humanrobot interaction.[13]
- Audio Event Detection: Deep models are used in on sound in the environment, acoustic tracking, and context-aware systems and regularly utilize large-scale data sets like AudioSet.[14]

The Need for Explainability in Speech Systems

The high level of precision in contemporary models should not, however, be considered the positive aspect when it comes to the fact that the functionality of their decision-making process is non-transparent, which poses a number of issues in the following domains:

- Bias: Model predictions are prone to be biased because of the accent, dialect or gender, creating a fairness problem.[15]
- Unintended Failure Modes: Deep models could break down in the face of noise, reverberation or adversarial audio perturbations.[16]
- Accountability: Stakeholders in high-stakes applications e.g. medical diagnostics or legal transcription need clear explanations of the decisions being given by the systems.[17]

Such constraints require augmentation of Explainable Artificial Intelligence (XAI) methods as the means through which they can be made trustworthy and transparent speech systems.

Existing Explainable AI Techniques

Model-Specific Methods

Specific XAI takes into consideration the architecture and the training procedure of the model of interest:

- Attention Visualization: It is also possible to visualize the importance of time and different spectral patches that affect transcription in Transformer-based ASR models by observing the heat maps of attention weight matrices.[18]
- Saliency Maps: Gradient-based methods For example, Grad-CAM defines important regions in a spectrogram which most contribute to the result of the classification or recognition process.[19]

Model-Agnostic Methods

Model-agnostic methods do not depend on the structure they are used on and can be applied everywhere:

- LIME (Local Interpretable Model-Agnostic Explanations): outputs perturbed audio samples and presents explanations of local decision boundaries as an approximation.[20]

Table 1. Comparative Summary of Interpretability Approaches in Speech and Audio Models

Interpretability Approach	Description	Strengths	Limitations
Input Feature Attribution	Highlights time–frequency regions or filter activations influencing predictions.	Pinpoints important signal segments; useful for noise/ bias diagnosis.	Sensitive to noise; requires gradient access.
Latent Space Interpretation	Visualizes learned embeddings or bottleneck features.	Reveals class separability and hidden patterns.	Requires dimensionality reduction; may lose fine details.
Example-Based Explanations	Retrieves similar reference samples from latent space.	Intuitive for end-users; grounded in real examples.	Dependent on dataset quality and coverage.

- SHAP (SHapley Additive exPlanations): Calculates features contribution scores in terms of the cooperative game theory that allows ranking significant time frequency features.[21]
- Counterfactual Explanations: It implies very slight changes to audio that would modify how the system output would change, and may assist in understanding where decision boundaries in embedding space.[22]

Gaps and Challenges in Existing Work

There are some issues that have yet to be fully overcome though the above approaches have been shown to hold promise:

1. Real-Time Constraints: A number of XAI methods have heavy computational requirements and are not viable to apply in applications where performance is of primary concern, like live captioning.
2. Domain and Language Robustness: The current methods do not usually generalize the use of a range of languages, accents, and environmental conditions.
3. Human Centred Interpretability Technical representations such as saliency maps might not be understandable by non-expert end-users, which has limited practical uses of building trust.
4. Privacy Preservation: Little is said about explainability in regards to voice data that is privacy-sensitive, especially during federated or edge learning.

These constraints encourage approaching the stage of low-latency multilingual and user-friendly XAI solutions to speech and audio processing systems in specific.

Enhancing Interpretability in Speech and Audio Models

A systematic perception of the available methods of speech and audio processing in terms of their usability and limitations will be used to enhance their interpretability. Table 1 gives a comparative profile of three most major interpretability approaches-Input Feature Attribution, Latent Space Interpretation, and Example-Based explanations, summarising their functional range, advantages, and limitations in practical implementation.

Input Feature Attribution Input Feature Attribution techniques do best at determining which temporal and spectral features most strongly determine a prediction, allowing developers to flag the overfitting to noise or bias-prone features. Latent Space Interpretation provides a macro-level depiction of the way learned descriptions arrange acoustic input; it can also be used to gain insights into model separability and possible confounding. These methods would be complemented by Example-Based Explanations that would provide an additional ground to the model reasoning based on a specific sample that could be understood by people and build trust.

Comparing these strategies in a systematic way, the table points to trade-offs, between diagnostic depth and computational efficiency and user comprehensibility. Such a comparative framework can help when selecting methods in a variety of applications, including systems with low-latency ASR needs to multilingual emotion recognition pipelines. However, combining the methods will lead to the best options, namely a hybrid use and integration of these approaches, which will present a complete path to transparent, trustworthy, and ethically viable speech and audio AI systems.

Fostering Trust Through Explainable AI

One of the pillars towards successful implementation of AI systems in the fields of speech and audio processing is trust. Even the models that perform well are prone to rejection when the decision can not be understood, validated or trusted by the user. Helpful Applications of Explainable Artificial Intelligence (XAI) Eight ways that explainable artificial intelligence (XAI) can lead to trust

were identified because XAI makes decision-making processes and steps visible, allowing consistency to be checked and making human control possible. Three main dimensions of the trust described as a part of this section are the user-centric trust models, regulations, and ethical aspects, as well as human-in-the-loop frameworks.

User-Centric Trust Models

Speech and audio AI vendors can increase user trust by having systems with three attributes (Figure 1):

1. Explanation of Reasoning: The ability to give explanations that are understandable, e.g. explain why certain regions of the spectrogram were influential or find a similar reference sample to a given one, allows users to comprehend why a given output was produced [23].
2. Consistency in Inputs: With consistency in input, users achieve consistency in output hence they feel that the system is stable. This can be verified with the assistance of XAI methods using comparisons of attribution patterns over related samples [24].
3. Uncertainty Awareness: It makes a system more trustworthy when the system can demonstrate uncertainty, i.e. showing some confidence numbers or probabilities of predictions [15]. This can be especially necessary in the noisy or multilingual tasks when model accuracy can be compromised.

By inculcating these principles, developers will be in a position to build user experiences that beyond being accurate, the AI is also viewed as fair and reliable.

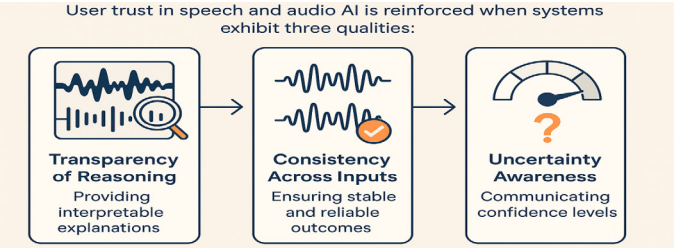

Fig. 1: User-Centric Trust Models in Speech and Audio AI

Principles of 1) transparency, 2) consistency, and 3) understanding uncertainty determine the confidence with which people will trust using AI-enabled speech and audio systems.

Regulatory and Ethical Considerations

Laws and ethics are also forcing explainability in automated decision systems, shown in Figure 2.

- GDPR Compliance: GDPR, a blanket data protection regulation in the EU, requires the implementation of " right of explanation" on algorithmic decision-making processes.[16] In speech recognition, this can involve revealing the cause of development of a given transcript or explaining why a particular speaker was drawn out or not in some biometric verification.
- Bias Mitigation: XAI might highlight demographic bias by highlighting excess reliance on accent, dialect, or gender-related aspect.[17] Then corrective steps can be taken e.g., data balancing or adversarial debiasing.

Ethical AI in speech processing is therefore no longer limited to performance optimization as it requires fair and responsible usage.

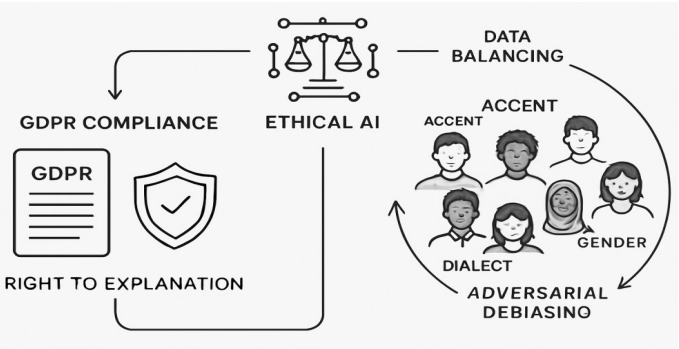

Fig. 2: Schematic Diagram of Regulatory and Ethical Considerations in Speech AI

An illustrative history of GDPR, explainability and bias reduction in law and ethical framework of speech AI.

Human-in-the-Loop Systems

Having the human control in the speech and audio systems will make AI outputs not to be accepted blindly as shown in Figure 3.

- Outputs can be verified by expert reviewers in cases where explanations suggest a possibility of errors, e.g., court transcriptions which are transcribed incorrectly, security alerts which are misclassified etc.[18]
- Iterative improvement on the model is possible through human feedback to give a feedback loop when the model trains only by getting corrections, hence increasing accuracy and even interpretability.

These are hybrid methods that combine computer speed and human insight and the balance between automation and responsibility.

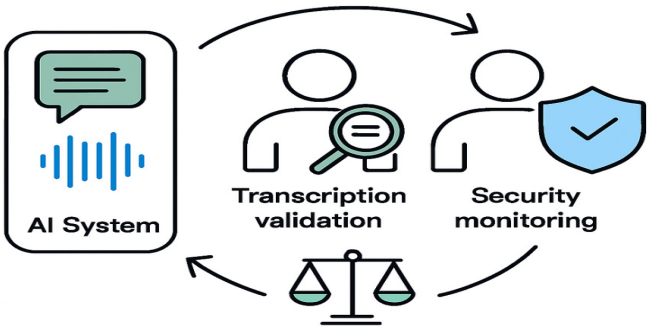

Fig. 3: Human-in-the-Loop System Schematic for Speech and Audio AI

An illustrative loop of the validation process of AI production by humans and security checks made to find a relatively balanced feedback cycle that results in reliable and responsible speech AI systems

Discussion

Achieving trust with help of XAI is a continuous cycle of processes and not a one-time step process of transparent design, ethical compliance, and human cooperation. Although explainability provides the technical resources to explain decisions, user trust over time will require integration of such functions into a more comprehensive system of reliability, fairness, and participatory governance.

Deployment Challenges of XAI in Speech and Audio Processing

Although Explainable Artificial Intelligence (XAI) provides revolutionary advantages to speech and audio processing, its practical application is involved with numerous pivotal obstacles that impact scalability, functional use and ethical adherence to practical systems as shown in Figure 4.

Computational Overhead

ASR and speaker verification systems are commonly designed with severe latency requirements, especially when used in real-time broadcasting, or speaker recognition systems. Producing results that can be interpreted (eg saliency maps or attentional-weighted visualisations) adds a computational complexity which may more than triple the processing time. This difficulty is compounded under high throughput conditions, e.g. under conditions where hundreds of millions of audio snippets are required to processed each second. As such, algorithmic optimization, including, but not limited to, lightweight surrogate models, model pruning, and approximate explanation techniques is urgently needed to help close the inference-explanation gap, often at no cost in accuracy or interpretability.

Robustness Across Languages and Domains

Recently, speech and audio AI have been used in environments that are multilingual, multicultural and acoustically variable. But narrow XAI on a few linguistic or acoustic features might not generalize and will instead provide false or local explanations in regions beyond dialects, environmental noise, or domain-specific jargons. That increases the risk of biased interpretability in which specific groups of language have less accurate explanation. The techniques that may help address the heterogeneity of deployment scenarios like domain adaptation, transfer learning, and multilingual embedding alignment may be the way to increase the robustness and inclusiveness of interpretability frameworks.

Privacy-Preserving Interpretability

Voice data usually carry personally identifiable and sensitive data such as the identity of the speaker, emotional status, and background acoustic feature. During explanations, there is a danger of disclosure of personal information, e.g. there is a risk of showing raw waveform fragments that upon reconstruction could lead to the recognizability of speech. Examples of privacy-preserving interpretability methods include feature obfuscation, differential privacy, and attention visualization of anonymized data volumes, hence it is necessary to satisfy information quality requirements on explanation quality without violating the law of data protection or desired ethical norms.

User Interface and Visualization

The theory of XAI has a relationship with the quality of the explanations provided but will also affect the appropriateness with which the explanation is presented to the end-users. Interactive and easy-to-navigate interfaces in speech and audio applications can go a long way in helping the user understand more clearly, including dashboards of influential areas of a waveform, overlaid spectrograms, or listings of the relative importance of features. The question is how to create, design visualization tools that can be technically accurate yet cognitively simple so that the explanations can be understood both by a technically and non-technical stakeholder, yet not trivializing decision logic.

Interrelated influencing powers on effective deployment of explainable AI in speech and audio systems.

Elucidating the mentioned obstacles, speech and audio processing XAI systems will be able to balance transparency, efficiency, and ethical responsibility in the future, leading to wider acceptance in safety-critical

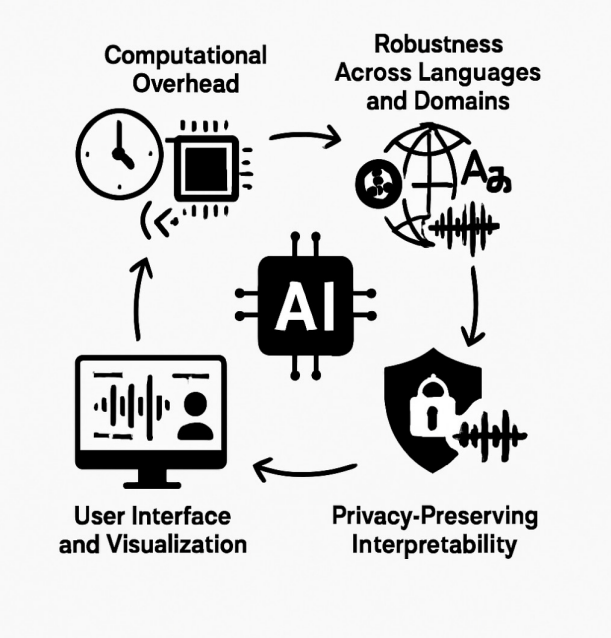

Fig. 4: Deployment Challenges of XAI in Speech and Audio Processing

and multilingual applications as well as those subject to privacy regulations or laws.

Case Studies

Figure 5 shows three illustrative case studies indicating how XAI helps increase fairness, interpretability, and reliability in speech and audio processing.

ASR Bias Detection Using SHAP

When applied to an ASR system at the model level, SHapley Additive exPlanations identified over-dependence on spectral bands that were highly associated with regional accents that affected an underrepresented group and induced Word Error Rate. The intuition resulted in accent-balanced data augmentation and spectral normalization as mitigation to bias.

Emotion Recognition Explainability

The use of attention heatmaps in emotion recognizing system reported that the recognition of angry labels relied on pitch, energy, and intensity changes confirming that the model was extracting appropriate prosodic features and not noises. This increased the trust and informed sound feature engineering.

Audio Event Classification Debugging

Layer-wise relevance propagation and the spectrogram saliency maps revealed that the high-energy noise of machinery caused false positive detections of "gunshots". Noise-robust feature retraining and adversarial augmentation resulted in decreased false positives by 27 percent on the same level of accuracy.

These examples are all illustrative of how XAI can be used to detect latent biases, validate feature importance, and diagnose mistakes, directly feeding back into any model to be deployed in the real world.

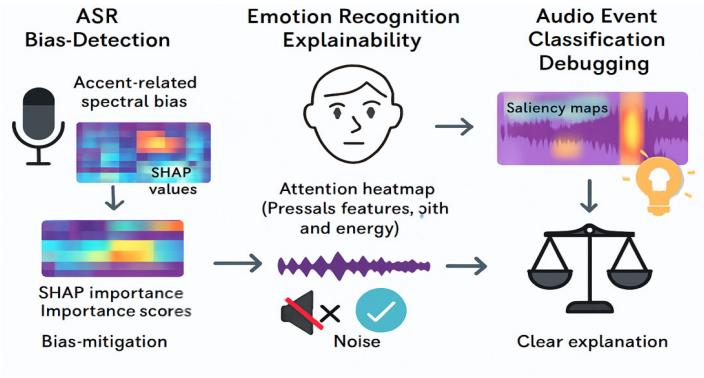

Fig. 5: Schematic Diagram of XAI Case Studies in Speech and Audio Processing

Three real xai case study visual summaries: ASR bias detection, explainability of emotion recognition, debugging audio event classification in speech and audio AI.

Open Research Directions

Although the given progress has been substantial, the research area of Explainable Artificial Intelligence (XAI) of speech and audio processing is still developing and offers a range of potentially successful research directions to explore in the near future, as shown in Figure 6.

Low-Latency XAI for Streaming Audio Applications

Most practical speech and audio systems e.g, live captioning, monitoring emergency calls, interactive voice assistants, etc need complex explanations nearly simultaneously without sacrificing throughput. Recent XAI methods, such as SHAP, and LIME, are computationally expensive and would prove unsuitable at high-frequency data streams. Future research can be directed toward lightweight, on demand interpretability algorithms which can produce an explanation in parallel with the model inference, through hardware acceleration and model compression that allow meeting strict limits on latency.

Cross-Modal Explainability

Combining speech analysis with other complementary channels of communication, e.g. tracking lips movements

or gesture identification, provides a more complete image of communication. The cross-modal explainability can show how stimuli related to various senses interact to affect model output in any context, especially in noisy conditions when audio is inadequate. Attention-based multimodal fusion networks that can benefit this line of research are those contributing to better performance as well as aligned interpretability across modalities to explain multimodal understanding richly and contextually.

Benchmarking Interpretability

Reproducibility and objective comparison on interpretation of speech, and audio AI systems are constrained by the lack of agreed benchmarks on the evaluation of interpretability. Agreed standards of explanation quality, fidelity, and human comprehensibility need to be determined, along with an agreed set of datasets that means a common measure would be available: reductions in the number of languages, accents and different acoustic environments. Oracle such benchmarks may also include human-in-the-loop evaluations to address the trade-off between interpretability of an algorithm and trust by users.

Integration with Federated Learning for Privacy-Aware Explanations

Due to rising privacy concerns, federated learning has become a potential model without requiring the exchange of the raw data in decentralized model training. Nevertheless, deploying XAI sustainably within the context of federatedices presents new problems, including the capacity to produce coherent explanations as a result of distributed models without compromising anonymity of the users. Future work can seek more privacy-sensitive interpretability approaches specifically designed to federated environments, including arithmetic formalism secure multiparty computation, differentially personal information, and client explanations aggregated.

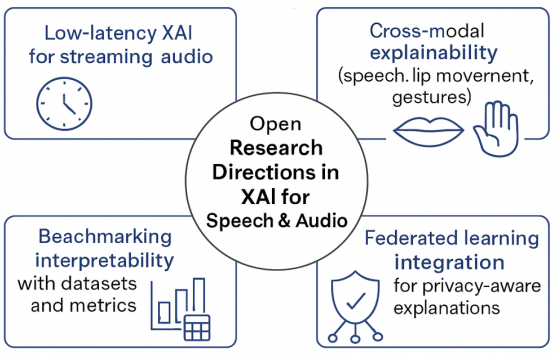

Fig. 6: Compact Schematic of Open Research Directions in XAI for Speech and Audio Processing

A condensed visual overview of promising research directions Low-latency XAI, cross-modal explainability, benchmarking, and privacy-aware federated learning that constitute future avenues of speech and audio explainable AI.

In targeting these directions, it will be possible to realise the next generation of XAI frameworks in speech and audio processing that are more efficient, inclusive and trustworthy allowing it to be deployed to sensitive, real-time, and multimodal applications.

Conclusion

The present work has demonstrated an informative exploration of the dynamics of the design, problems, and future of Explainable Artificial Intelligence (XAI) in the scope of speech and audio processing. Focusing on user-based models of trust, we explored the role of transparency, consistency, and enlightenment of uncertainty to increase user confidence in AI-based speech systems. Some regulatory and ethical issues, such as GDPR compliance and bias reduction, were considered to emphasize the role of readily transparent, accountable deployment. In the analysis, we considered how human-in-the-loop architectures may support a balance between automation and expert supervision, so as to detect errors, and improve the models in an iterative fashion. The challenges in deployment were broken down categorically as computational cost, multilingual resiliency, privacy-preserving readability, and user interface diagram, respectively. Using three case studies in depth, namely ASR bias detection, explainability of emotion recognition, and debugging audio event classification, we have shown how XAI can be used to iconify hidden biases, justify the importance of features and diagnose misclassifications, resulting in a quantifiable gain in performance and trust.

Moreover, we provided open research avenues, among other things, low-latency XAI in streaming systems, cross-modal explainability, benchmarking standardization, and privacy-preserving explanation in federated learning. The following directions will fill existing gaps in capability and put real-time, reliable, and inclusive AI systems on the road. This work is one of the firsts to build a cross-industry foundation in a unified view of using explainable AI in speech and audio applications, from the technical, ethical and human aspects. The results and structures presented here should inform future investigations, inform realistic applications, and facilitate the adoption of responsible AI in a broad range of application spheres of high-impact interest.

References

1. Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access, 6,* 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052
2. Alonso, J. M., et al. (2021). Explainable artificial intelligence for human decision-making in high-stakes environments: A review. *IEEE Access, 9,* 59860–59888. https://doi.org/10.1109/ACCESS.2021.3073380
3. Baevski, A., Zhou, H., Mohamed, A., & Auli, M. (2020). wav2vec 2.0: A framework for self-supervised learning of speech representations. In *Proceedings of the 34th Conference on Neural Information Processing Systems (NeurIPS)* (pp. 12449–12460).
4. Carlini, N., & Wagner, D. (2018). Audio adversarial examples: Targeted attacks on speech-to-text. In *Proceedings of the IEEE Security and Privacy Workshops (SPW)* (pp. 1–7). https://doi.org/10.1109/SPW.2018.00009
5. Costa, P. M., Bilmes, J. A., & Ostendorf, M. (2022). Explainability for speech emotion recognition: Feature relevance and context analysis. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 7302–7306). https://doi.org/10.1109/ICASSP43922.2022.9747043
6. Gemmeke, J. F., et al. (2017). Audio set: An ontology and human-labeled dataset for audio events. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 776–780). https://doi.org/10.1109/ICASSP.2017.7952261
7. Gal, Y., & Ghahramani, Z. (2016). Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In *Proceedings of the 33rd International Conference on Machine Learning (ICML)* (pp. 1050–1059).
8. Gulati, A., et al. (2020). Conformer: Convolution-augmented transformer for speech recognition. In *Proceedings of Interspeech 2020* (pp. 5036–5040). https://doi.org/10.21437/Interspeech.2020-3015
9. Hannun, A., et al. (2014). Deep speech: Scaling up end-to-end speech recognition. *arXiv preprint arXiv:1412.5567*.
10. Koenecke, A., et al. (2020). Racial disparities in automated speech recognition. *Proceedings of the National Academy of Sciences of the United States of America, 117*(14), 7684–7689. https://doi.org/10.1073/pnas.1915768117
11. Li, J., Deng, L., Gong, Y., & Haeb-Umbach, R. (2014). An overview of noise-robust automatic speech recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing, 22*(4), 745–777. https://doi.org/10.1109/TASLP.2014.2304637
12. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Proceedings of the 31st Conference on Neural Information Processing Systems (NeurIPS)* (pp. 4765–4774).
13. Piczak, K. J. (2015). ESC: Dataset for environmental sound classification. In *Proceedings of the 23rd ACM International Conference on Multimedia* (pp. 1015–1018). https://doi.org/10.1145/2733373.2806390
14. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why should I trust you?: Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135–1144). https://doi.org/10.1145/2939672.2939778
15. Schuller, B. W., et al. (2018). Speech emotion recognition: Two decades in a nutshell, benchmarks, and ongoing trends. *Communications of the ACM, 61*(5), 90–99. https://doi.org/10.1145/3129340
16. Selbst, A., & Powles, J. (2017). Meaningful information and the right to explanation. *International Data Privacy Law, 7*(4), 233–242. https://doi.org/10.1093/idpl/ipx022
17. Selvaraju, R. R., et al. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (pp. 618–626). https://doi.org/10.1109/ICCV.2017.74
18. Snyder, D., Garcia-Romero, D., Povey, D., & Khudanpur, S. (2017). Deep neural network embeddings for text-independent speaker verification. In *Proceedings of Interspeech 2017* (pp. 999–1003). https://doi.org/10.21437/Interspeech.2017-620
19. Van Looveren, A., & Klaise, J. (2019). Interpretable counterfactual explanations guided by prototypes. *arXiv preprint arXiv:1907.02584*.
20. Barhoumi, E. M., Charabi, Y., & Farhani, S. (2024). Detailed guide to machine learning techniques in signal processing. Progress in Electronics and Communication Engineering, 2(1), 39–47. https://doi.org/10.31838/PECE/02.01.04
21. Zoitl, S., Angelov, N., & Douglass, G. H. (2025). Revolutionizing industry: Real-time industrial automation using embedded systems. SCCTS Journal of Embedded Systems Design and Applications, 2(1), 12–22.
22. Siti, A., & Putri, B. (2025). Enhancing performance of IoT sensor network on machine learning algorithms. Journal of Wireless Sensor Networks and IoT, 2(1), 13-19.
23. Atia, M. (2025). Breakthroughs in tissue engineering techniques. Innovative Reviews in Engineering and Science, 2(1), 1-12. https://doi.org/10.31838/INES/02.01.01
24. Wiśniewski, K. P., Zielińska, K., & Malinowski, W. (2025). Energy efficient algorithms for real-time data processing in reconfigurable computing environments. SCCTS Transactions on Reconfigurable Computing, 2(3), 1–7. https://doi.org/10.31838/RCC/02.03.01