Novelty Assessment Report

Paper: EmotionThinker: Prosody-Aware Reinforcement Learning for Explainable Speech Emotion Reasoning
PDF URL: https://openreview.net/pdf?id=wbttgzp7MT
Venue: ICLR 2026 Conference Submission
Year: 2026
Report Generated: 2026-01-05

Abstract

Emotional information in speech plays a unique role in multimodal perception. However, current Speech Large Language Models (SpeechLLMs), similar to conventional speech emotion recognition (SER) systems, still treat emotion understanding as a simple classification problem. This provides limited interpretability of predictions, while leaving the LLMs' expressive and reasoning capabilities underutilized. In this work, we take the first step to reformulate SER as a deep reasoning problem through reinforcement learning (RL). We propose EmotionThinker, which is designed to generate accurate emotion predictions with interpretable explanations grounded in fine-grained acoustic cues. To achieve this, we first construct EmotionCoT-35K, an emotional reasoning dataset with Chain-of-Thought annotations and detailed captions. Second, we observe that current SpeechLLMs exhibit weak prosody perception, whereas prosodic cues constitute fundamental signals for interpreting emotions. To address this, we develop the prosody-enhanced foundation model EmotionThinker-Base, and demonstrate that prosody enhancement improves emotion understanding. Third, we introduce Group-Relative-Policy-Optimization with Progressive-Trust-aware-Reasoning-Reward (GRPO-PTR}) for RL. Different from standard GRPO, which relies only on rule-based outcome rewards, GRPO-PTR progressively introduces reasoning reward, dynamically adjusts it with a trustworthiness weight reflecting the alignment between reasoning and outcome, and evaluates the overall reasoning quality with a reward model based on multi-dimensional criteria. EmotionThinker outperforms previous state-of-the-art evaluation models both in emotion accuracy and explanation quality, advancing SER toward interpretable multimodal reasoning.

Core Task Landscape

This paper addresses: **explainable speech emotion recognition with reasoning**

A total of **50 papers** were analyzed and organized into a taxonomy with **20 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Explainability Techniques and Interpretability Methods**
- **Reasoning-Based Approaches**
- **Multimodal Integration**
- **Speech-Only Emotion Recognition**
- **Conversational and Contextual Emotion Understanding**
- **Cross-Lingual Emotion Detection Tasks**
- **Affective Computing Foundations and Applications**

Complete Taxonomy Tree

- explainable speech emotion recognition with reasoning Survey Taxonomy
- Explainability Techniques and Interpretability Methods
  - Feature-Level Explainability (4 papers)
  - [5] Speech emotion recognition using overlapping sliding window and Shapley additive explainable deep neural network (Nhat Truong Pham, 2023) View paper
  - [8] Unveiling hidden factors: explainable AI for feature boosting in speech emotion recognition (Nfissi, 2024) View paper
  - [15] Informative Speech Features based on Emotion Classes and Gender in Explainable Speech Emotion Recognition (HÃ¼seyin YÄ±ldÄ±rÄ±m, 2023) View paper
  - [25] Iterative Feature Boosting for Explainable Speech Emotion Recognition (Nfissi, 2023) View paper
  - Attention-Based Temporal Explainability (2 papers)
  - [10] Evaluating significant features in contextâ□aware multimodal emotion recognition with XAI methods (Aaishwarya Khalane, 2025) View paper
  - [42] Explainable speech emotion recognition through attentive pooling: insights from attention-based temporal localization (Bouchigny, 2025) View paper
  - Model-Agnostic Explanation Methods (2 papers)
  - [44] Exploring Local Interpretable Model-Agnostic Explanations for Speech Emotion Recognition with Distribution-Shift (MJ Hjuler, 2025) View paper
  - [50] Speech Emotion Recognition with Explainable AI: Enhancing Transparency in Emotion Recognition Systems (Sejal Sharma, 2025) View paper
  - Multimodal Explainability Frameworks (6 papers)
  - [1] Multimodal emotion recognition with explainable AI for cognitive human-computer interaction in smart environments (Sarah S, 2025) View paper
  - [2] Speech emotion recognition using deep learning transfer models and explainable techniques (Tae-Wan Kim, 2024) View paper
  - [11] Explainable artificial intelligence techniques for speech emotion recognition: A focus on xai models (Michael Norval, 2025) View paper

    - [32] Bridging Text and Speech for Emotion Understanding: An Explainable Multimodal Transformer Fusion Framework with Unified Audioâ□□Text Attribution (Ashutosh Pandey, 2025) View paper
    - [35] Towards the Explainability of Multimodal Speech Emotion Recognition (Puneet Kumar, 2021) View paper
    - [49] Interpretable multimodal emotion recognition using hybrid fusion of speech and image data (Puneet Kumar, 2023) View paper
- Reasoning-Based Approaches
    - Chain-of-Thought and Generative Reasoning (2 papers)
    - [16] Think out Loud: Emotion Deducing Explanation in Dialogues (Li Jiang-nan, 2024) View paper
    - [20] Towards a Generative Approach for Emotion Detection and Reasoning (Strzalkowski, 2024) View paper
    - Reinforcement Learning for Reasoning ★ (3 papers)
    - [0] EmotionThinker: Prosody-Aware Reinforcement Learning for Explainable Speech Emotion Reasoning (Anon et al., 2026) View paper
    - [6] R1-omni: Explainable omni-multimodal emotion recognition with reinforcement learning (Zhao, 2025) View paper
    - [37] EMO-RL: Emotion-Rule-Based Reinforcement Learning Enhanced Audio-Language Model for Generalized Speech Emotion Recognition (Pengcheng Li, 2025) View paper
    - Multimodal Reasoning Frameworks (5 papers)
    - [19] Explainable Multimodal Emotion Reasoning (Lian Zheng, 2023) View paper
    - [36] Reasoning Beyond Majority Vote: An Explainable SpeechLM Framework for Speech Emotion Recognition (Su Bo-Hao, 2025) View paper
    - [40] Emotion-LLaMA: Multimodal Emotion Recognition and Reasoning with Instruction Tuning (Zebang Cheng, 2024) View paper
    - [48] Beyond Classification: Towards Speech Emotion Reasoning with Multitask AudioLLMs (Zhang Wen-yu, 2025) View paper
- Multimodal Integration
    - Transformer-Based Multimodal Fusion (3 papers)
    - [4] Interpretable multimodal emotion recognition using optimized transformer model with SHAP-based transparency (Adel A. Alyoubi, 2025) View paper
    - [22] Multimodal attentive learning for real-time explainable emotion recognition in conversations (Balaji Arumugam, 2022) View paper
    - [31] Enhancing Emotion Detection Accuracy and Transparency Through Multimodal Fusion and Explainable AI (S.Keerthiga, 2025) View paper
    - Other Multimodal Architectures (2 papers)
    - [26] Emotion recognition using explainable genetically optimized fuzzy ART ensembles (W. S. Liew, 2021) View paper
    - [33] Explainable cross-domain emotion recognition using non-linear optimization and multimodal feature fusion based deep learning model (Baazeem, 2025) View paper
- Speech-Only Emotion Recognition
    - Deep Learning and Transfer Learning for SER (1 papers)
    - [46] Speech Emotion Recognition in Continuous Space by Using Machine Learning (Uluocak, 2019) View paper
    - Federated and Privacy-Preserving SER (2 papers)
    - [30] FedSER-XAI: PSO-optimized multi-stream cross-attention transformer with graph features for explainable federated speech emotion recognition. (Eman Abdulrahman Alkhamali, 2025) View paper
    - [41] DEEMO: De-identity Multimodal Emotion Recognition and Reasoning (Li Deng, 2025) View paper
    - Cross-Lingual and Multilingual Emotion Recognition (1 papers)
    - [45] Bilingual Speech Emotion Recognition Using Neural Networks: A Case Study for Turkish and English Languages (Damla BÃ¼Å□ra Ã□zsÃ¶nmez, 2021) View paper
- Conversational and Contextual Emotion Understanding
    - Emotion Flip Detection and Reasoning (3 papers)
    - [13] YNU-HPCC at SemEval-2024 Task10: Pre-trained Language Model for Emotion Discovery and Reasoning its Flip in Conversation (Chenyi Liang, 2024) View paper
    - [24] COMMA-DEER: COmmon-sense Aware Multimodal Multitask Approach for Detection of Emotion and Emotional Reasoning in Conversations (Bhattacharyya, 2022) View paper
    - [39] MultiFlipFormer: A Multimodal Transformer for Emotion Flip Reasoning and Instigator Detection in Therapeutic Conversations (A Sharma, 2025) View paper
    - Debiasing and Fairness in Emotion Recognition (1 papers)
    - [23] A Training-Free Debiasing Framework with Counterfactual Reasoning for Conversational Emotion Detection (Jing Ran, 2023) View paper
- Cross-Lingual Emotion Detection Tasks
    - Shared Task Systems and Benchmarks (3 papers)
    - [7] Findings of the WASSA 2024 EXALT shared task on Explainability for Cross-Lingual Emotion in Tweets (Aaron Maladry, 2024) View paper
    - [14] HITSZ-HLT at WASSA-2024 Shared Task 2: Language-agnostic Multi-task Learning for Explainability of Cross-lingual Emotion Detection (Wang Jun, 2024) View paper
    - [17] WU_TLAXE at WASSA 2024 Explainability for Cross-Lingual Emotion in Tweets Shared Task 1: Emotion through Translation using TwHIN-BERT and GPT (Levow, 2024) View paper
    - Cross-Lingual Transfer and Ensemble Methods (1 papers)
    - [18] TEII: Think, Explain, Interact and Iterate with Large Language Models to Solve Cross-lingual Emotion Detection (Bi Sheng, 2024) View paper
    - Code-Mixed Conversational Emotion Tasks (1 papers)
    - [9] IASBS at SemEval-2024 Task 10: Delving into Emotion Discovery and Reasoning in Code-Mixed Conversations (Mehrzad Tareh, 2024) View paper
- Affective Computing Foundations and Applications
    - Surveys and Reviews on Explainable Affective Computing (2 papers)
    - [12] Toward explainable affective computing: A review (Karina CortiÃ±as-Lorenzo, 2023) View paper
    - [28] Explainable Artificial Intelligence (XAI) for Emotion Detection (Madan Mohan Tito Ayyalasomayajula, 2024) View paper
    - Application-Oriented Emotion Recognition Systems (6 papers)
    - [21] Toward explainable AI (XAI) for mental health detection based on language behavior (Kerz, 2023) View paper
    - [29] Emotion-Infused Models for Explainable Psychological Stress Detection (McKeown, 2021) View paper

- [34] Ai-driven emotion recognition for mental health diagnoses: Assessing mental health through emotional state evaluation (PRETO, 2025) View paper
- [38] Causality Aware Multimodal Reasoning Network in Human Emotion Identification and Sentiment Understanding. (NK Thakre, 2025) View paper
- [43] Designing Explainable In-vehicle Interfaces for Conditionally Automated Driving: A Holistic Examination with Mixed Method Approaches (WANG, 2024) View paper
- [47] Human-Centered AI: Designing Emotion-Aware Systems for Safe Humanâ□□Machine Interaction (Chakravarty, 2022) View paper
- Unrelated Domain Applications (1 papers)
- [3] Revolutionizing Prostate Cancer Diagnosis: Vision Transformers with Explainable Artificial Intelligence to Accurate and Interpretable Prostate Cancer Identification (Krunal Maheriya, 2025) View paper

Narrative

Core task: explainable speech emotion recognition with reasoning. The field organizes around several complementary branches that address different facets of understanding and justifying emotion predictions from speech. Explainability Techniques and Interpretability Methods focus on making model decisions transparent through attention mechanisms, saliency maps, and post-hoc analysis tools such as LIME and SHAP, as seen in works like XAI Speech Techniques[11] and Distribution-Shift LIME[44]. Reasoning-Based Approaches emphasize structured inference and logical chains, often leveraging reinforcement learning or chain-of-thought prompting to produce human-understandable rationales. Multimodal Integration combines acoustic signals with text, video, or physiological data to enrich emotion understanding, while Speech-Only Emotion Recognition concentrates on purely acoustic feature extraction and modeling. Conversational and Contextual Emotion Understanding examines dialogue history and speaker interactions, and Cross-Lingual Emotion Detection Tasks explore generalization across languages. Finally, Affective Computing Foundations and Applications ground the technical work in real-world scenarios such as mental health monitoring and human-computer interaction.

Within Reasoning-Based Approaches, a small but growing cluster explores reinforcement learning to guide models toward interpretable decision paths. EmotionThinker[0] exemplifies this direction by training agents to generate step-by-step reasoning traces that justify emotion labels, aligning closely with EMO-RL[37] and R1-omni[6], which similarly apply RL frameworks to refine reasoning quality. These methods contrast with purely attention-driven explainability (e.g., Multimodal Explainable Emotion[1]) by explicitly optimizing for coherent rationales rather than relying on post-hoc visualization. A key trade-off is computational cost versus interpretability depth: RL-based reasoning can yield richer explanations but requires careful reward design and longer training. EmotionThinker[0] sits at the intersection of reasoning and explainability, offering a middle ground where the model learns to articulate its logic during inference, bridging the gap between black-box performance and human-centered transparency.

Related Works in Same Category

The following **2 sibling papers** share the same taxonomy leaf node with the original paper:

1. R1-omni: Explainable omni-multimodal emotion recognition with reinforcement learning

Authors: Zhao, Jiaxing, Wei, Xihan, Bo, et al. (6 authors total) | **Year/Venue**: 2025 | **URL**: View paper

Abstract

In this work, we present the first application of Reinforcement Learning with Verifiable Reward (RLVR) to an Omni-multimodal large language model in the context of emotion recognition, a task where both visual and audio modalities play crucial roles. We leverage RLVR to optimize the Omni model, significantly enhancing its performance in three key aspects: reasoning capability, emotion recognition accuracy, and generalization ability. The introduction of RLVR not only improves the model's overall...

Relationship Analysis

Both papers belong to the 'Reinforcement Learning for Reasoning' category, employing RL to optimize emotion reasoning quality and explanation generation in speech emotion recognition systems. They overlap in using RL-based approaches (GRPO) to enhance explainability through chain-of-thought reasoning and verifiable rewards for emotion classification accuracy. However, EmotionThinker focuses specifically on prosody-aware reasoning with a novel GRPO-PTR framework that introduces progressive trust-aware reasoning rewards and operates on speech-only inputs, while R1-Omni applies RLVR to an omni-multimodal model (video + audio) for emotion recognition without the prosody-enhancement stage or trust-aware reward mechanisms.

2. EMO-RL: Emotion-Rule-Based Reinforcement Learning Enhanced Audio-Language Model for Generalized Speech Emotion Recognition

Authors: Pengcheng Li, Botao Zhao, Zuheng Kang, Junqing Peng, Qu Xiaoyang, et al. (7 authors total) | **Year/Venue**: 2025 | **URL**: View paper

Abstract

â□¦ enhancing explainability via cognitive transparency â□¦ of different reasoning strategies in EMO-RL: implicit reasoning, â□¦ that address the unique challenges of speech emotion detection. â□¦

Relationship Analysis

Both papers belong to the Reinforcement Learning for Reasoning category, employing RL to optimize emotion reasoning quality in speech emotion recognition systems. They overlap in using GRPO-based RL frameworks to enhance explainable SER with reasoning capabilities, both building on Qwen2-Audio architectures and incorporating structured reasoning formats. The key differences are that EmotionThinker focuses on prosody-aware reasoning with progressive trust-aware rewards (GRPO-PTR) and constructs EmotionCoT-35K with detailed prosodic annotations, while EMO-RL emphasizes emotion similarity-weighted rewards (ESWR) and explicit structured reasoning (ESR) to address convergence instability from ambiguous emotional boundaries.

Contributions Analysis

Overall novelty summary. The paper introduces EmotionThinker, which reformulates speech emotion recognition as a reasoning problem using reinforcement learning to generate interpretable explanations grounded in acoustic cues. It resides in the 'Reinforcement Learning for Reasoning' leaf under 'Reasoning-Based Approaches', alongside only two sibling papers (EMO-RL and R1-omni). This leaf represents a sparse, emerging research direction within the broader taxonomy of 50 papers across 36 topics, indicating that RL-driven reasoning for emotion understanding remains relatively unexplored compared to attention-based explainability or multimodal fusion approaches.

The taxonomy reveals that neighboring leaves focus on Chain-of-Thought generative reasoning (without RL) and Multimodal Reasoning Frameworks, while the parent branch 'Reasoning-Based Approaches' contrasts with 'Explainability Techniques' that emphasize post-hoc analysis tools like SHAP and LIME. EmotionThinker bridges reasoning generation and explainability by training models to articulate logic during inference, diverging from purely attention-driven methods in adjacent branches. The sparse population of its leaf suggests this RL-for-reasoning direction is less crowded than feature-level explainability or transformer-based multimodal fusion, which contain four to six papers each.

Among 20 candidates examined across three contributions, none were found to clearly refute the paper's claims. The EmotionCoT-35K dataset examined 3 candidates with 0 refutable; the prosody-enhanced foundation model examined 10 candidates with 0 refutable; and the GRPO-PTR framework examined 7 candidates with 0 refutable. This limited search scope—top-K semantic matches plus citation expansion—suggests that within the examined literature, no prior work directly overlaps with the combination of prosody enhancement, CoT annotations, and RL-based reasoning rewards. However, the small candidate pool means the analysis cannot confirm exhaustive novelty across the entire field.

Based on the 20-candidate search, the work appears to occupy a relatively novel position at the intersection of RL-driven reasoning and prosody-aware emotion understanding. The sparse taxonomy leaf and absence of refutable candidates within the examined scope suggest incremental but meaningful differentiation from existing methods. A broader literature search or deeper examination of the two sibling papers' technical details would be needed to assess whether the prosody enhancement and trust-aware reward mechanisms constitute substantial advances over prior RL-based reasoning frameworks.

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: EmotionCoT-35K dataset with prosody-aware Chain-of-Thought annotations

Description: The authors curate a training dataset of 35,000 speech–reasoning pairs spanning about 200 hours of audio, annotated with emotion labels and fine-grained prosodic features (pitch, energy, speed, stress, intonation) plus step-wise reasoning traces. This dataset enables models to produce both emotion labels and perceptually grounded explanations.

This contribution was assessed against **3 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Chain-of-Thought Distillation with Fine-Grained Acoustic Cues for Speech Emotion Recognition

URL: View paper

Brief Assessment

Chain-of-Thought Distillation[58] focuses on distilling reasoning abilities from large LLMs to domain-specific SER models using fine-grained acoustic features, but does not describe creating a large-scale prosody-annotated dataset like EmotionCoT-35K with 35,000 speech-reasoning pairs spanning 200 hours.

---

2. UniSS: Unified Expressive Speech-to-Speech Translation with Your Voice

URL: View paper

Brief Assessment

UniSS Expressive Translation[60] focuses on speech-to-speech translation with a different dataset (UniST, 44.8k hours) for cross-lingual translation tasks, not emotion recognition with prosodic reasoning annotations. The datasets serve fundamentally different purposes and application domains.

---

3. Plug-and-Play Emotion Graphs for Compositional Prompting in Zero-Shot Speech Emotion Recognition

URL: View paper

Brief Assessment

Plug-and-Play Emotion[59] focuses on zero-shot prompting with structured emotion graphs extracted via DSP tools, not on creating a training dataset with Chain-of-Thought annotations for supervised learning or RL.

Contribution 2: Prosody-enhanced foundation model EmotionThinker-Base

Description: The authors build a foundation model via prosody-centric supervised fine-tuning on approximately 500 hours of data, including stress perception, prosodic attribute classification, and comparative prosodic augmentation tasks. This stage equips the model with strong prosody perception ability before reinforcement learning.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Improving speech depression detection using transfer learning with wav2vec 2.0 in low-resource environments

URL: View paper

Brief Assessment

Depression Detection Transfer[62] focuses on transfer learning with wav2vec 2.0 for depression detection in speech, not on building prosody-enhanced foundation models for emotion understanding. The candidate addresses a different clinical application (depression) rather than general emotion reasoning with prosodic augmentation tasks.

---

2. EmoDubber: Towards High Quality and Emotion Controllable Movie Dubbing

URL: View paper

Brief Assessment

EmoDubber[64] focuses on movie dubbing with emotion control through flow-based synthesis, not on building prosody-enhanced foundation models for emotion understanding via supervised fine-tuning on prosody perception tasks.

---

3. Hierarchical Emotion Prediction and Control in Text-to-Speech Synthesis

URL: View paper

Brief Assessment

Hierarchical Emotion Control[68] focuses on text-to-speech synthesis with emotion control at different granularity levels (phoneme, word, utterance), not on building a prosody-enhanced foundation model for speech emotion recognition through reinforcement learning as in the original paper.

---

4. Prosody-Enhanced Acoustic Pre-training and Acoustic-Disentangled Prosody Adapting for Movie Dubbing

URL: View paper

Brief Assessment

Prosody-Enhanced Dubbing[63] focuses on movie dubbing tasks with acoustic-prosody disentanglement for video-synchronized speech generation, not on building a prosody-enhanced foundation model for emotion understanding via reinforcement learning as in the original paper.

---

5. ProsodyLM: Uncovering the Emerging Prosody Processing Capabilities in Speech Language Models

URL: View paper

Brief Assessment

ProsodyLM[66] focuses on pre-training speech language models with prosody-aware tokenization for general prosody processing capabilities, not on building emotion-specific foundation models through supervised fine-tuning with stress perception and emotion reasoning tasks as in the original paper.

---

6. PROEMO: Prompt-Driven Text-to-Speech Synthesis Based on Emotion and Intensity Control

URL: View paper

Brief Assessment

PROEMO[69] focuses on text-to-speech synthesis with emotion control through prompt-based methods and intensity encoders, not on building prosody-enhanced foundation models for speech emotion recognition via reinforcement learning. The technical domains and objectives differ fundamentally.

---

7. EmoSRE: Emotion prediction based speech synthesis and refined speech recognition using large language model and prosody encoding

URL: View paper

Brief Assessment

EmoSRE Prosody Encoding[61] focuses on speech synthesis and recognition using prosody encoding for emotion prediction, not on building a prosody-enhanced foundation model via supervised fine-tuning for emotion understanding as described in the original paper.

---

8. Cross Corpus Speech Emotion Recognition using transfer learning and attention-based fusion of Wav2Vec2 and prosody features

URL: View paper

Brief Assessment

Cross Corpus Wav2Vec2[65] focuses on cross-corpus speech emotion recognition using Wav2Vec2 features combined with prosody features via attention-based fusion. This is a different technical approach from building a foundation model through prosody-centric supervised fine-tuning on stress perception and prosodic attribute classification tasks as described in the original paper.

---

9. Disentangling Prosody Representations With Unsupervised Speech Reconstruction

URL: View paper

Brief Assessment

Disentangling Prosody[70] focuses on unsupervised prosody disentanglement for speech reconstruction and emotion recognition, not on building prosody-enhanced foundation models through supervised fine-tuning with stress perception and prosodic attribute classification tasks as described in the original paper.

---

10. Emotion-Aware Prosodic Phrasing for Expressive Text-to-Speech

URL: View paper

Brief Assessment

Emotion-Aware Prosodic[67] focuses on prosodic phrasing for text-to-speech synthesis, not on building prosody-enhanced foundation models for emotion understanding via reinforcement learning. The candidate addresses phrase break prediction in TTS, while the original develops a foundation model for speech emotion recognition with prosody-centric supervised fine-tuning.

Contribution 3: GRPO-PTR reinforcement learning framework with progressive trust-aware reasoning reward

Description: The authors propose a reinforcement learning strategy that progressively introduces a reasoning reward model trained on multi-dimensional criteria (factual alignment, interpretative quality, caption completeness, fluency) and dynamically adjusts it with a trustworthiness weight reflecting reasoning-outcome alignment. This approach supervises intermediate reasoning quality and mitigates reward hacking.

This contribution was assessed against **7 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Trusted Artificial Intelligence in Life-Critical and High-Risk Environments

URL: View paper

Brief Assessment

Trusted AI Life-Critical[56] focuses on integrating explainable statistical models into RL for minefield traversal and safety-critical autonomous systems, not on progressive reasoning reward mechanisms for emotion recognition or interpretable predictions in multimodal contexts.

---

2. Developing and Integrating Trust Modeling into Multi-Objective Reinforcement Learning for Intelligent Agricultural Management

URL: View paper

Brief Assessment

Trust Agricultural Management[51] focuses on integrating a trust model into multi-objective RL for agricultural fertilization management, not on progressive reasoning reward mechanisms for interpretable predictions in speech emotion recognition.

---

3. Modeling Trust and Deception in Multi-Agent Reinforcement Learning Using the Werewolf Game

URL: View paper

Brief Assessment

Werewolf Game Trust[54] focuses on trust modeling in multi-agent game environments with symbolic decision logic, not on progressive reasoning reward mechanisms for interpretable predictions in speech emotion recognition tasks.

---

4. Implementation of Human-AI Interaction in Reinforcement Learning: Literature Review and Case Studies

URL: View paper

Brief Assessment

Human-AI Interaction Review[57] is a survey paper focused on human-AI interaction mechanisms in reinforcement learning, covering learning from human feedback, demonstrations, shared autonomy, querying, and explainability. It does not propose novel RL training algorithms or progressive reward mechanisms for reasoning tasks.

---

5. Rethinking Subjective Trust in LLM: Actualizing Tangibility from Uncertainty

URL: View paper

Brief Assessment

Subjective Trust LLM[55] focuses on calibrated trust and defensive training in LLMs through alignment methods, not on progressive trust-aware reasoning rewards for speech emotion recognition with prosodic supervision.

---

6. Socio-Cognitive Recommendation via Neuroscience-Inspired Decision Dynamics and Cognitive-Social Signal Fusion

URL: View paper

Brief Assessment

Neuroscience-Inspired Recommendation[53] focuses on recommendation systems using neuroscience-inspired decision dynamics and social signal fusion, not on reinforcement learning frameworks for speech emotion reasoning with progressive trust-aware rewards.

---

7. NSCTI: A Hybrid Neuro-Symbolic Framework for AI-Driven Predictive Cyber Threat Intelligence

URL: View paper

Brief Assessment

NSCTI Cyber Threat[52] focuses on cybersecurity threat detection using neuro-symbolic learning and federated intelligence, not on reinforcement learning for explainable emotion reasoning or progressive trust-aware reasoning rewards for speech models.

Appendix: Text Similarity Detection

No high-similarity text segments were detected across any compared papers.

References

- [0] EmotionThinker: Prosody-Aware Reinforcement Learning for Explainable Speech Emotion Reasoning View paper
- [1] Multimodal emotion recognition with explainable AI for cognitive human-computer interaction in smart environments View paper
- [2] Speech emotion recognition using deep learning transfer models and explainable techniques View paper
- [3] Revolutionizing Prostate Cancer Diagnosis: Vision Transformers with Explainable Artificial Intelligence to Accurate and Interpretable Prostate Cancer Identification View paper
- [4] Interpretable multimodal emotion recognition using optimized transformer model with SHAP-based transparency View paper
- [5] Speech emotion recognition using overlapping sliding window and Shapley additive explainable deep neural network View paper
- [6] R1-omni: Explainable omni-multimodal emotion recognition with reinforcement learning View paper
- [7] Findings of the WASSA 2024 EXALT shared task on Explainability for Cross-Lingual Emotion in Tweets View paper
- [8] Unveiling hidden factors: explainable AI for feature boosting in speech emotion recognition View paper
- [9] IASBS at SemEval-2024 Task 10: Delving into Emotion Discovery and Reasoning in Code-Mixed Conversations View paper
- [10] Evaluating significant features in contextâ□aware multimodal emotion recognition with XAI methods View paper
- [11] Explainable artificial intelligence techniques for speech emotion recognition: A focus on xai models View paper
- [12] Toward explainable affective computing: A review View paper
- [13] YNU-HPCC at SemEval-2024 Task10: Pre-trained Language Model for Emotion Discovery and Reasoning its Flip in Conversation View paper
- [14] HITSZ-HLT at WASSA-2024 Shared Task 2: Language-agnostic Multi-task Learning for Explainability of Cross-lingual Emotion Detection View paper
- [15] Informative Speech Features based on Emotion Classes and Gender in Explainable Speech Emotion Recognition View paper
- [16] Think out Loud: Emotion Deducing Explanation in Dialogues View paper
- [17] WU_TLAXE at WASSA 2024 Explainability for Cross-Lingual Emotion in Tweets Shared Task 1: Emotion through Translation using TwHIN-BERT and GPT View paper
- [18] TEII: Think, Explain, Interact and Iterate with Large Language Models to Solve Cross-lingual Emotion Detection View paper
- [19] Explainable Multimodal Emotion Reasoning View paper
- [20] Towards a Generative Approach for Emotion Detection and Reasoning View paper
- [21] Toward explainable AI (XAI) for mental health detection based on language behavior View paper
- [22] Multimodal attentive learning for real-time explainable emotion recognition in conversations View paper
- [23] A Training-Free Debiasing Framework with Counterfactual Reasoning for Conversational Emotion Detection View paper
- [24] COMMA-DEER: COmmon-sense Aware Multimodal Multitask Approach for Detection of Emotion and Emotional Reasoning in Conversations View paper
- [25] Iterative Feature Boosting for Explainable Speech Emotion Recognition View paper
- [26] Emotion recognition using explainable genetically optimized fuzzy ART ensembles View paper
- [27] Explainable multimodal emotion recognition View paper
- [28] Explainable Artificial Intelligence (XAI) for Emotion Detection View paper
- [29] Emotion-Infused Models for Explainable Psychological Stress Detection View paper
- [30] FedSER-XAI: PSO-optimized multi-stream cross-attention transformer with graph features for explainable federated speech emotion recognition. View paper
- [31] Enhancing Emotion Detection Accuracy and Transparency Through Multimodal Fusion and Explainable AI View paper
- [32] Bridging Text and Speech for Emotion Understanding: An Explainable Multimodal Transformer Fusion Framework with Unified Audioâ□Text Attribution View paper
- [33] Explainable cross-domain emotion recognition using non-linear optimization and multimodal feature fusion based deep learning model View paper
- [34] Ai-driven emotion recognition for mental health diagnoses: Assessing mental health through emotional state evaluation View paper
- [35] Towards the Explainability of Multimodal Speech Emotion Recognition View paper
- [36] Reasoning Beyond Majority Vote: An Explainable SpeechLM Framework for Speech Emotion Recognition View paper
- [37] EMO-RL: Emotion-Rule-Based Reinforcement Learning Enhanced Audio-Language Model for Generalized Speech Emotion Recognition View paper
- [38] Causality Aware Multimodal Reasoning Network in Human Emotion Identification and Sentiment Understanding. View paper
- [39] MultiFlipFormer: A Multimodal Transformer for Emotion Flip Reasoning and Instigator Detection in Therapeutic Conversations View paper

- [40] Emotion-LLaMA: Multimodal Emotion Recognition and Reasoning with Instruction Tuning View paper
- [41] DEEMO: De-identity Multimodal Emotion Recognition and Reasoning View paper
- [42] Explainable speech emotion recognition through attentive pooling: insights from attention-based temporal localization View paper
- [43] Designing Explainable In-vehicle Interfaces for Conditionally Automated Driving: A Holistic Examination with Mixed Method Approaches View paper
- [44] Exploring Local Interpretable Model-Agnostic Explanations for Speech Emotion Recognition with Distribution-Shift View paper
- [45] Bilingual Speech Emotion Recognition Using Neural Networks: A Case Study for Turkish and English Languages View paper
- [46] Speech Emotion Recognition in Continuous Space by Using Machine Learning View paper
- [47] Human-Centered AI: Designing Emotion-Aware Systems for Safe Humanâ□□Machine Interaction View paper
- [48] Beyond Classification: Towards Speech Emotion Reasoning with Multitask AudioLLMs View paper
- [49] Interpretable multimodal emotion recognition using hybrid fusion of speech and image data View paper
- [50] Speech Emotion Recognition with Explainable AI: Enhancing Transparency in Emotion Recognition Systems View paper
- [51] Developing and Integrating Trust Modeling into Multi-Objective Reinforcement Learning for Intelligent Agricultural Management View paper
- [52] NSCTI: A Hybrid Neuro-Symbolic Framework for AI-Driven Predictive Cyber Threat Intelligence View paper
- [53] Socio-Cognitive Recommendation via Neuroscience-Inspired Decision Dynamics and Cognitive-Social Signal Fusion View paper
- [54] Modeling Trust and Deception in Multi-Agent Reinforcement Learning Using the Werewolf Game View paper
- [55] Rethinking Subjective Trust in LLM: Actualizing Tangibility from Uncertainty View paper
- [56] Trusted Artificial Intelligence in Life-Critical and High-Risk Environments View paper
- [57] Implementation of Human-AI Interaction in Reinforcement Learning: Literature Review and Case Studies View paper
- [58] Chain-of-Thought Distillation with Fine-Grained Acoustic Cues for Speech Emotion Recognition View paper
- [59] Plug-and-Play Emotion Graphs for Compositional Prompting in Zero-Shot Speech Emotion Recognition View paper
- [60] UniSS: Unified Expressive Speech-to-Speech Translation with Your Voice View paper
- [61] EmoSRE: Emotion prediction based speech synthesis and refined speech recognition using large language model and prosody encoding View paper
- [62] Improving speech depression detection using transfer learning with wav2vec 2.0 in low-resource environments View paper
- [63] Prosody-Enhanced Acoustic Pre-training and Acoustic-Disentangled Prosody Adapting for Movie Dubbing View paper
- [64] EmoDubber: Towards High Quality and Emotion Controllable Movie Dubbing View paper
- [65] Cross Corpus Speech Emotion Recognition using transfer learning and attention-based fusion of Wav2Vec2 and prosody features View paper
- [66] ProsodyLM: Uncovering the Emerging Prosody Processing Capabilities in Speech Language Models View paper
- [67] Emotion-Aware Prosodic Phrasing for Expressive Text-to-Speech View paper
- [68] Hierarchical Emotion Prediction and Control in Text-to-Speech Synthesis View paper
- [69] PROEMO: Prompt-Driven Text-to-Speech Synthesis Based on Emotion and Intensity Control View paper
- [70] Disentangling Prosody Representations With Unsupervised Speech Reconstruction View paper