

# How does the scaling of OpenPangu-7B-MLA’s performance on EchoMind compare to prosody-exclusive models when tr

Assignee Research

June 10, 2026

## Abstract

Speech Language Models (SLMs) have made significant progress in spoken language understanding. Yet it remains unclear whether they can fully perceive non lexical vocal cues alongside spoken words, and respond with empathy that aligns with both emotional and contextual factors. Existing benchmarks typically evaluate linguistic, acoustic, reasoning, or dialogue abilities in isolation, overlooking the integration of these skills that is crucial for human-like, emotionally intelligent conversation. We present EchoMind, the first interrelated, multi-level benchmark that simulates the cognitive proc

## 1 Introduction

This paper examines: EchoMind: An Interrelated Multi-level Benchmark for Evaluating Empathetic Speech Language Models. Research question: How does the scaling of OpenPangu-7B-MLA’s performance on EchoMind compare to prosody-exclusive models when trained on varying amounts of low-resource language data, measured by cross-entropy loss and perplexity?.

## 2 Methodology

Systematic literature search across multiple databases yielded 11 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.5/10.

## 3 Results

11 papers retrieved. 12 claims extracted; 1 independently verified. Quality review score: 4.5/10.

## 4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

## 5 Extracted Claims

Claim	Verified	Confidence
EchoMind is an interrelated multi-level benchmark for evaluating empathetic speech language models.	✓	0.32
EchoMind evaluates both text and audio inputs and outputs.	×	0.03
EchoMind includes tasks for understanding, reasoning, conversation, content, and voice.	×	0.08
EchoMind supports multi-level evaluation (M).	×	0.08
EchoMind includes tasks for speaker, paragraph, and environment understanding.	×	0.03
EchoMind includes tasks for reasoning, conversation, content, and voice.	×	0.07
EchoMind includes tasks for style understanding.	×	0.07
EchoMind includes tasks for response (audio) evaluation.	×	0.04
EchoMind includes tasks for NISQA, DNMOs, EmoAlign, and VES evaluation.	×	0.02
EchoMind includes tasks for text context fit, text response naturalness, text colloquial degree, and text speech relevance	×	0.07
EchoMind includes tasks for audio VES and audio quality evaluation.	×	0.03
EchoMind includes tasks for context fit, speech relevance, and VES evaluation.	×	0.04

## References

- <http://arxiv.org/abs/2412.10008v1>
- <http://arxiv.org/abs/2510.22758v2>
- <http://arxiv.org/abs/2106.09063v4>