大语言模型辅助用于可检查的多模态情绪数据标注的工具包

ZHEYUAN KUANG，悉尼大学，澳大利亚WEIWEI JIANG，中国南京信息工程大学NICHOLAS KOEMEL，Charles Perkins Centre，医学与健康学院，澳大利亚；Turner脑与心理健康研究所，莫纳什大学，澳大利亚MATTHEW AHMADI，Charles Perkins Centre，医学与健康学院，澳大利亚；Turner脑与心理健康研究所，莫纳什大学，澳大利亚EMMANUEL STAMATAKIS，Charles Perkins Centre，医学与健康学院，澳大利亚；Turner脑与心理健康研究所，莫纳什大学，澳大利亚BENJAMIN TAG，新南威尔士大学，澳大利亚ANUSHA WITHANA，悉尼大学，澳大利亚ZHANNA SARSENBAYEVA，悉尼大学，澳大利亚

arXiv:2603.02569v1 [cs.HC] 3 Mar 2026

多模态情绪识别（MER）日益依赖细粒度、以证据为依据的标注，但当线索在不同模态间动态变化且彼此错位时，检查和标签构建便难以规模化。我们提出一种大语言模型辅助工具包，通过可检查、以事件为中心的工作流程支持多模态情绪数据标注。该工具包对异构录制数据进行预处理和对齐，在交互式共享时间轴上可视化所有模态，并将结构化信号呈现为视频轨道，以进行跨模态一致性检查。随后，它检测候选事件，并将同步关键帧和时间窗口打包为事件数据包，同时附上可追溯至源数据的指针。最后，该工具包集成大语言模型、特定模态工具和提示模板，起草结构化标注，供分析人员验证和编辑。我们通过具有代表性的示例，展示了该工作流程在多模态虚拟现实情绪录制数据上的应用。

CCS 概念：• **以人为中心的计算** → **交互系统与工具**。

其他关键词和短语：人机协作标注、大语言模型、多模态情绪识别

1 引言

多模态情绪识别（MER）面临着诸多实际挑战，而这些挑战往往源于数据处理工作，而非模型设计。情绪相关线索分布在不同模态中，并随时间变化，因此跨模态整合和事件级解读十分困难。这提高了对细粒度、描述性且以证据为依据的标注的需求。然而，基准测试结果表明，标注成本高、已标注数据有限，以及在领域差异、噪声和模态缺失情况下鲁棒性下降，仍然是主要障碍 [8, 9]。这些限制使可扩展的数据检查和一致的标签构建成为 MER 面临的核心挑战。

作者联系信息：ZHEYUAN KUANG，悉尼大学，澳大利亚悉尼，zheyuan.kuang@sydney.edu.au；WEIWEI JIANG，南京信息工程大学，中国南京，weiweijiangcn@gmail.com；NICHOLAS KOEMEL，Charles Perkins Centre，医学与健康学院，澳大利亚悉尼；以及 Turner脑与心理健康研究所，莫纳什大学，澳大利亚墨尔本，nicholas.koemel@ sydney.edu.au；MATTHEW AHMADI，Charles Perkins Centre，医学与健康学院，澳大利亚悉尼；以及 Turner脑与心理健康研究所，莫纳什大学，澳大利亚墨尔本，matthew.ahmadi@sydney.edu.au；EMMANUEL STAMATAKIS，Charles Perkins Centre，医学与健康学院，澳大利亚悉尼；以及 Turner脑与心理健康研究所，莫纳什大学，澳大利亚墨尔本，emmanuel.stamatakis@sydney. edu.au；BENJAMIN TAG，新南威尔士大学，澳大利亚悉尼，benjamin.tag@unsw.edu.au；ANUSHA WITHANA，悉尼大学，澳大利亚悉尼，anusha.withana@sydney.edu.au；ZHANNA SARSENBAYEVA，悉尼大学，澳大利亚悉尼，zhanna.sarsenbayeva@sydney.edu.au。

针对这些挑战，近期的 MER 研究开始利用大语言模型，以支持更具描述性和开放性的情绪理解 [8]。在实际应用中，许多 MER 研究的一个局限在于数据检查：分析人员需要手动对齐多个视频和信号图，以追踪不同模态中动态且往往错位的线索，并判断哪些时间窗口应视为候选事件 [13]。在此类数据处理工作中，大语言模型可提供互补能力，以一致的格式生成结构化多模态标注，减少人工标注工作量，同时支持分析人员核验和编辑 [14, 15]。

我们通过开发一款大语言模型辅助工具包[1] 来弥补这些不足，该工具包将多模态情绪数据标注的三阶段工作流程付诸实践。该工具包首先对异构录制数据进行预处理和对齐，然后在交互式共享时间轴上可视化所有模态，并将结构化信号呈现为视频轨道，以支持可检查的跨模态一致性检查。其次，它检测候选事件，并检索相应的关键帧和时间窗口，将其整理为带有可追溯指针的事件数据包。第三，它利用配备特定模态工具和提示模板的大语言模型起草结构化标注，供分析人员核验和编辑。这些阶段共同支持从多模态情绪数据集中大规模构建可用于训练的标签。我们通过具有代表性的示例展示了该工作流程在多模态虚拟现实情绪记录上的应用。

2 相关工作

近期研究开始将大语言模型纳入数据工作流程，让大语言模型参与预处理和检查，但仍需人工核验，并明确记录转换过程 [6]。研究人员探索将大语言模型用作数据预处理器，以执行错误检测、插补和匹配等任务，并指出了促使人们谨慎检查的实际局限性 [15]。可视分析领域的研究也认为，大语言模型能够在交互式检查过程中帮助人们浏览、查询和总结复杂数据，同时强调依据锚定和风险管理，而非盲目自动化 [4]。在标注领域，人类与大语言模型协作工作流程表明，大语言模型能够提出标签和解释，由人类核验不确定的案例，从而提升效率而不削弱人工控制 [14]。

大语言模型还可以通过生成候选标签来支持情绪数据集构建，并通过人工检查或模型验证来管理质量。例如，Jing 等人 [5] 使用 GPT-4o 和结构化提示对多模态情绪数据集进行标注，并通过人工审查和下游训练验证标签，展示了其效用以及验证的必要性。Niu 等人 [11] 指出，大语言模型的情绪判断可能与人工标注系统性地存在差异，并主张主要将大语言模型用于分流和质量控制，而非完全取代人工。

3 提出的方法

我们提出了一款大语言模型辅助工具包，可将多模态情绪数据标注的数据处理工作流程转化为可操作的三阶段流程。该工具包将异构录制数据转换为可供检查、以事件为中心的输出，支持直观检查信号变化、进行情绪相关事件检测和关键帧检索，以及生成基于证据的标注。它支持多种数据模态，包括但不限于视频流和结构化时间序列信号。

我们在来自84 名参与者、涵盖六种情绪诱发场景的多模态虚拟现实情绪记录上演示该工具包，其中包括基于虚拟化身的面部表情视频、第一人称视角视频、结构化身体动作、血容量脉搏（BVP）、心率（HR）、皮肤电活动（EDA）以及惯性测量单元信号。如图 1所示，该界面提供同步的多模态轨道、事件时间轴和标注视图，用于以事件为中心的检查与标注。

[1]论文被接收后，工具包代码仓库将以开源形式发布。

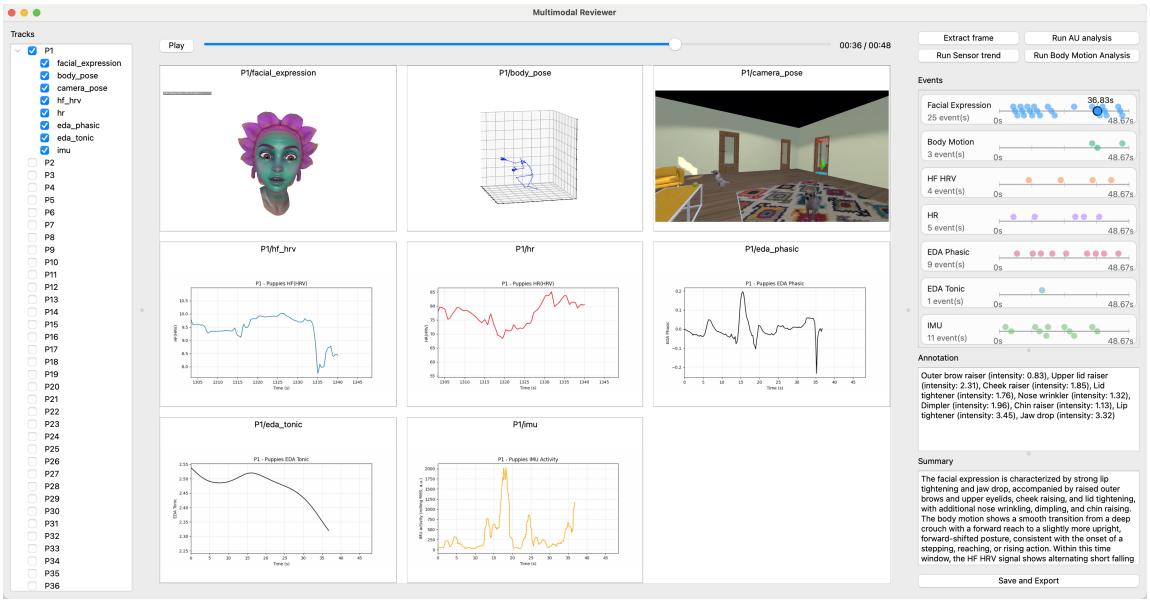

图1. 用于检查和标注多模态情绪数据的工具包界面。

3.1 预处理与可视化

在第一阶段，该工具包执行预处理与统一可视化。其目标是使异构模态能够在共享时间线上被联合检查，以便分析人员高效观察信号变化并定位潜在事件。我们首先使用一致的会话元数据对每种模态进行标准化，包括参与者标识符、VR场景标识符、采样率和时间戳。对于结构化信号，我们将每条数据流渲染为视频轨道，从而使时间序列能够以与视频相同的交互范式进行检查。这使得在单一界面中对不同模态轨道进行时间锁定回放和并排比较成为可能。

随后，我们将所有轨道同步到统一的时间轴，并将对齐结果保存为索引，以支持跨模态导航、缩放以及对任意时间窗口的可复现检索。生成的用户界面按参与者和 VR 场景组织会话，提供概览以便快速浏览，并提供详细视图以检查细粒度的时间模式、行为和生理信号峰值。此阶段输出对齐后的可视化轨道和会话级索引，为事件挖掘和后续标注奠定基础。

3.2 事件检测与检索

在第二阶段，工具包会检测信号变化事件，并检索各模态中对应的关键帧和时间窗口。对于面部表情视频，我们使用动作单元（AU） [3] 峰值检测，并借助 OpenFace [1], 将每个峰值视为候选事件。对于身体运动信号，我们计算运动能量指标，以识别运动事件 [12]。对于生理信号，我们提取峰值窗口和趋势变化窗口，以捕捉候选事件。所有候选事件都会与统一时间轴对齐，并汇总到用户界面的事件层中，使分析人员能够跳转到事件、比较跨模态上下文并细化事件边界。我们将此层呈现为交互式事件时间线，点击事件标记即可跳转到其时间戳，并显示各模态中对齐的帧。

分析人员可以在用户界面中验证、编辑或舍弃候选事件。随后，工具包会将每个事件打包为一个事件包，其中包含指向原始数据的可追溯指针，包括所有模态的关键帧和时间窗口。

3.3 大语言模型注释

在第三阶段，工具包支持对提取出的事件进行大语言模型辅助情绪标注。此阶段的目标是将每个事件窗口转换为标准化标注和证据摘要，用于情绪相关事件检测，并为下游情绪识别模型构建训练数据。

我们设计了一套特定于模态的预处理工具和提示模板，引导大语言模型生成字段一致的结构化标注。我们使用 GPT-5.2 作为后端大语言模型，以利用其多模态推理能力。对于面部表情，我们将峰值动作单元帧映射为简短的文本描述 [2]。对于身体运动，大语言模型会生成标注，以简洁的姿势和运动线索描述事件窗口内的骨架序列 [10]。对于生理信号，大语言模型会使用趋势和峰值描述符总结窗口内容，包括上升或下降片段、局部极值和持续时间 [7]。对于第一人称视角流，大语言模型会提取活动和环境线索等上下文信息，以支持对事件的解读。最后，大语言模型会汇总单模态描述，为每个事件生成详细的多模态描述，从而进一步完善标注 [2], ，并将其作为额外的情绪描述符添加到事件包中。

分析人员在用户界面中检查生成的标注，跳转至引用的时间范围进行验证，并编辑或舍弃错误字段。最终确定的标注随后会导出为结构化记录，供下游使用。

4 结论与未来工作

本文介绍了一款大语言模型辅助工具包，用于可检查、以事件为中心的多模态情绪数据标注。该工具包会对异构录制数据进行预处理，并在交互式共享时间轴上对齐；将结构化信号呈现为视频轨道，并支持候选事件检索。对于每个事件，大语言模型都会生成标注，供分析人员验证和编辑，从而支持更具扩展性且一致性更高的多模态情绪识别数据处理工作。

未来，我们将通过邀请领域专家参与用户研究来评估该工具包，并将检查效率和标注结果与标准实践进行比较。我们还计划在更多数据集上验证工作流程，并将工具包扩展至更多模态以及更灵活的标注模式。

参考文献

[1] Tadas Baltrušaitis、Peter Robinson 和 Louis-Philippe Morency。2016。Openface：一款开源面部行为分析工具包。载于*WACV*。1–10。[2] Zebang Cheng、Zhi-Qi Cheng、Jun-Yan He、Kai Wang、Yuxiang Lin、Zheng Lian、Xiaojiang Peng 和 Alexander Hauptmann。2024。Emotion-llama：通过指令微调实现多模态情绪识别与推理。神经信息处理系统进展 37（2024），110805–110853。[3] Paul Ekman 和 Wallace V Friesen。1978。面部动作编码系统。环境心理学与非语言行为 （1978）。[4] Maeve Hutchinson、Radu Jianu、Aidan Slingsby 和 Pranava Madhyastha。2024。大语言模型辅助的视觉分析：机遇与挑战。*arXiv:2409.02691*（2024）。[5] Xin Jing、Jiadong Wang、Iosif Tsangko、Andreas Triantafyllopoulos 和 Björn W Schuller。2025。MELT：利用大语言模型嵌入知识实现自动化多模态情绪数据标注。*arXiv:2505.24493*（2025）。[6] Hong Jin Kang、Muhammad Ali Gulzar、Nanyun Peng、Miryung Kim 等。2024。通过来源追踪实现人在回路中的合成文本数据检查。载于计算语言学协会论文集：*NAACL 2024*年。3118–3129。[7] Zechen Li、Shohreh Deldari、Linyao Chen、Hao Xue 和 Flora D Salim。2025。Sensorllm：使大语言模型与运动传感器对齐，以实现人类活动识别。载于*2025*年自然语言处理实证方法会议论文集。354–379。[8] Zheng Lian、Rui Liu、Kele Xu、Bin Liu、Xuefei Liu、Yazhou Zhang、Xin Liu、Yong Li、Zebang Cheng、Haolin Zuo 等。2025。MER 2025：当情感计算遇上大语言模型。载于第*33*届*ACM*国际多媒体会议论文集。13837–13842。[9] Zheng Lian、Licai Sun、Yong Ren、Hao Gu、Haiyang Sun、Lan Chen、Bin Liu 和 Jianhua Tao。2026。MERBench：多模态情绪识别统一评估基准。*IEEE*模式分析与机器智能汇刊（2026）。

[10] Haifeng Lu、Jiuyi Chen、Feng Liang、Mingkui Tan、Runhao Zeng 和 Xiping Hu。2025。借助大语言模型理解情绪性身体表达。载于*AAAI*人工智能大会论文集，第39卷。1447–1455。[11] Minxue Niu、Yara El-Tawil、Amrit Romana 和 Emily Mower Provost。2025。在大语言模型时代重新思考情绪标注。*IEEE*情感计算汇刊 （2025）。[12] Fotini Patrona、Anargyros Chatzitofis、Dimitrios Zarpalas 和 Petros Daras。2018。运动分析：基于动作捕捉数据的动作检测、识别与评估。模式识别 76（2018），612–622。[13] Nikki A Puccetti、William J Villano、Jonathan P Fadok 和 Aaron S Heller。2022。大脑中情感的时间动态：来自人体成像和动物模型的证据。《神经科学与生物行为评论》133（2022），104491。[14] Xinru Wang、Hannah Kim、Sajjadur Rahman、Kushan Mitra 和 Zhengjie Miao。2024。通过有效验证大语言模型标签实现人机协作标注。载于*2024*年*CHI*计算机系统中的人为因素大会论文集。1–21。[15] Haochen Zhang、Yuyang Dong、Chuan Xiao 和 Masafumi Oyamada。2023。将大语言模型用作数据预处理器。*arXiv:2308.16361* （2023）。