

Generative Multimodal Analysis (GMA) for Learning Process Data Analytics

Ridwan Whitehead¹, Andy Nguyen¹ and Sanna Järvelä¹

¹ Learning and Educational Technology (LET) Research Lab, University of Oulu, Pentti Kaiteran katu 1, Oulu, Finland

Abstract

This paper introduces Generative Multimodal Analysis (GMA), a novel method designed for utilizing Artificial Intelligence (GenAI) in the analysis of multimodal data derived from learning processes. The method is encapsulated in a systematic framework that integrates and optimizes GenAI technology with multimodal large language models (MLLMs) for application in multimodal learning analytics. The recent emergence and advancement of GenAI, particularly MLLMs, has opened new avenues for the automated interpretation and meaningful analysis of varied data sources. Current research in the field has sightseen diverse applications of GenAI in transforming learning and teaching practice. However, there is a noticeable gap in systematic methodologies for applying GenAI to scrutinize learning process data. This paper aims to bridge this gap by proposing the GMA method in the sphere of multimodal learning analytics with learning process data. In addition to the proposed methodological framework, this study also proposes an operational prototype for the practical implementation of GMA. This prototype serves as a tool for examining multimodal data in learning processes. To demonstrate the applicability and effectiveness of our proposed method, we conducted and presented a case study. Our approach offers essential guidance for learning scientists and educational technology application developers, reflecting the contemporary trends and needs in educational technologies. By providing a structured, innovative approach for employing GenAI in learning process data analysis, this study contributes significantly to the advancement of learning analytics methods.

Keywords

Generative Artificial Intelligence; Learning Analytics; Multimodal Data

1. Introduction

The rapid evolution of educational technologies and learning sciences, particularly the groundbreaking strides in artificial intelligence (AI) technology, has heralded an era where data-driven insights have become valuable in enhancing learning processes. This intersection of AI with educational methodologies is reshaping how information is analyzed, interpreted, and applied to improve teaching strategies and learning outcomes. For instance, Luckin and Cukurova [1] highlighted the potential of AI to provide personalized learning experiences, adapting to individual learner's needs. Likewise, Holmes et al. [2] discussed the transformative role of AI in education, particularly in providing insights into learning patterns. Recently, Järvelä et al. [3] proposed a human-AI collaboration approach for better unfolding the learning processes in the context of socially shared regulation of learning. Despite these advancements, the application of AI in learning analytics has primarily been fragmented. Current research predominantly focuses on isolated applications of AI in educational contexts [4], [5], lacking a comprehensive methodology for systematic analysis. This gap is particularly apparent in the sphere of generative Artificial Intelligence (GenAI), where its potential for learning process data analytics remains underexplored.

The advent of GenAI and its integration into educational contexts, particularly through multimodal large language models (MLLMs), represents a significant leap in the domain of educational technology. GenAI, a specialized subset of artificial intelligence that focuses on generating new content, has seen rapid evolution over the past decade. A key development in this evolution has been the creation of sophisticated multimodal large language models (MLLMs), such as OpenAI's GPT (Generative Pre-trained Transformer) series, which exemplify the advancements in natural language processing and understanding. The release of ChatGPT, based on the GPT architecture, in November 2022 represented a significant milestone [6], [7]. ChatGPT, with its advanced language understanding and generation

Joint Proceedings of LAK 2024 Workshops, co-located with 14th International Conference on Learning Analytics and Knowledge (LAK 2024), Kyoto, Japan, March 18-22, 2024.

✉ ridwan.whitehead@oulu.fi (R. Whitehead); andy.nguyen@oulu.fi (A. Nguyen); sanna.jarvela@oulu.fi (S. Järvelä)

ORCID 0009-0002-2888-7304 (R. Whitehead); 0000-0002-0759-9656 (A. Nguyen); 0000-0001-6223-3668 (S. Järvelä)



© 2023 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

capabilities, offered a more interactive and intuitive way for users to engage with AI. Furthermore, GenAI opens new possibilities in content creation and data analysis. Accordingly, learning analytics, encompassing the collection, measurement, analysis, and reporting of data about learners and their contexts, could greatly benefit from the integration of GenAI. This integration enables more nuanced extraction of insights from diverse learning process data, thereby enhancing our understanding of student learning behaviors and outcomes.

This paper addresses the identified gap in comprehensive methodologies for GenAI in multimodal learning analytics and proposes Generative Multimodal Analysis (GMA) as a structured approach to employ GenAI effectively in this field. Such a methodology is essential due to the growing complexity of learning environments and the variety of data they produce. As argued by Järvelä and Bannert [8], the integration of multimodal data analysis in educational research is crucial for a holistic understanding of learning processes. As an integral part of this study, we have developed a specialized software tool designed for GMA. This tool is intended to serve as a practical resource for researchers and educational developers, providing them with a robust and user-friendly platform to apply GMA methodologies in their work. The effectiveness and relevance of this software tool are demonstrated through its application in a case study. This implementation not only showcases the tool's functionality but also affirms GMA's efficacy and broad applicability in the field of educational research and development.

2. Generative Multimodal Analysis (GMA) for Learning Process Data Analytics

Generative Multimodal Analysis (GMA) represents a comprehensive methodological framework designed to transform the approach of researchers and analysts in educational settings. This framework leverages the capabilities of GenAI to expedite and enrich the process of extracting and integrating verbal and nonverbal elements from learning process data. Furthermore, GMA extends its integration to include material objects and the environment, capturing how they are utilized by learners engaged in active interaction within the learning context. By harnessing the power of GenAI, GMA facilitates a more profound and holistic understanding of the multifaceted nature of learning environments, where verbal communication, nonverbal cues, physical objects, and the surrounding environment all play integral roles in the learning process.

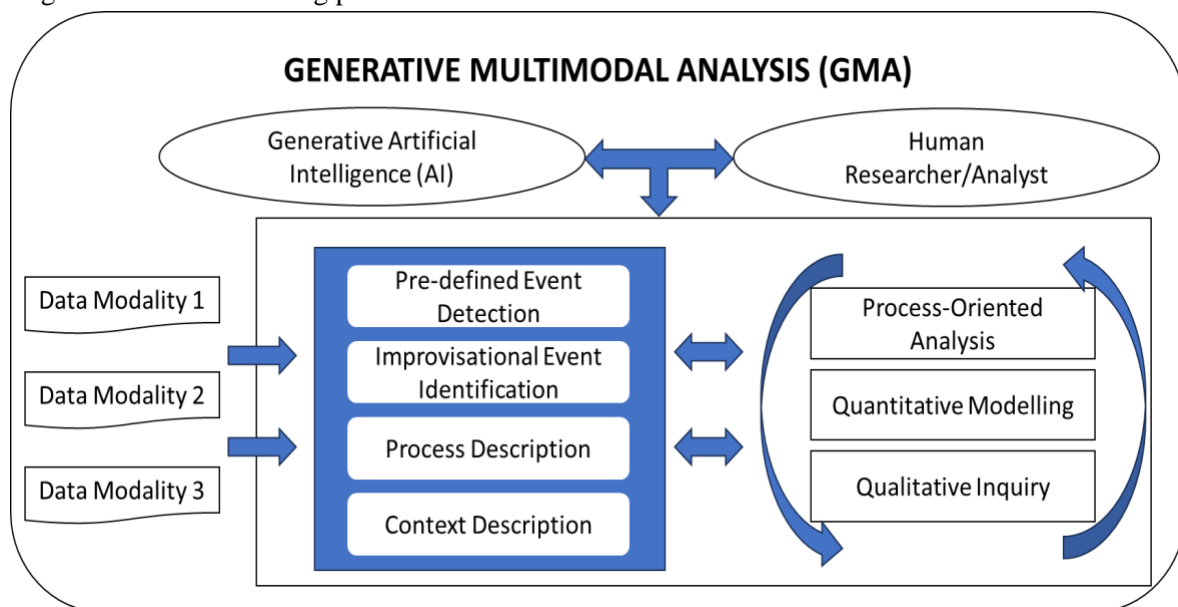


Figure 1: Generative Multimodal analysis (GMA) Framework

Figure 1 provides a visual depiction of the Generative Multimodal Analysis (GMA) Framework, illustrating its dynamic capabilities in leveraging generative AI for educational research. Within this framework, generative AI is adept at producing various types of outputs, including 1) predefined events, 2) improvisational events, 3) detailed descriptions of the learning process, and 4) comprehensive descriptions of the learning context. Crucially, as per the human-AI collaboration approach in research suggested by Järvelä, Nguyen and Hadwin [3], it's imperative that these AI-generated outputs are

subjected to evaluation and validation by human researchers or analysts. This collaborative approach ensures that the insights offered by AI are grounded in human understanding and expertise as well as its reliability.

Furthermore, the information generated through the GMA Framework can be dissected and examined through various analytical lenses. These include a) process-oriented analysis, which focuses on the dynamics and phases of the learning process; b) quantitative modeling, offering a statistical perspective and uncovering patterns and correlations; and c) qualitative inquiry, which delves into the deeper, nuanced aspects of the learning environment and experiences. This multifaceted analytical approach allows for a comprehensive and multi-dimensional understanding of the learning process, harnessing the strengths of both AI and human analysis.

As for illustration, Figure 2 provides a demonstration of how the GMA is applied to automatically extract specific pre-defined events. In this case, the focus is on identifying and analysing the non-verbal posture states of learners engaged in a collaborative learning setting. The figure showcases the capability of GMA to discern and categorize various non-verbal cues, particularly the postures of learners, which play a significant role in understanding engagement, interaction dynamics, and overall effectiveness of collaborative learning processes. This example highlights the advanced analytical power of GMA in recognizing and interpreting subtle yet crucial aspects of learner behaviour in collaborative learning.

Figure 3 showcases the user interface of our newly developed Generative Multimodal Analysis (GMA)

frame_27726_right



```
{
  "id": "chatcmpl-877yP1f24ZVPu1191XDxh8LvIyH",
  "object": "chat.completion",
  "created": 1701951267,
  "model": "gpt-4-1106-vision-preview",
  "usage": {
    "prompt_tokens": 1203,
    "completion_tokens": 17,
    "total_tokens": 1220
  },
  "choices": [
    {
      "message": {
        "role": "assistant",
        "content": "Open posture because the person is leaning forward with uncrossed arms, indicating engagement."
      },
      "finish_details": {
        "type": "stop",
        "stop": "<|fim_suffix|>"
      },
      "index": 0
    }
  ]
}
```

Figure 2: Example of GMA for detecting pre-defined events

Toolkit for describing video data. This interface is specifically designed to facilitate direct interaction with video data, particularly focusing on observational data from learning processes.

The example presented within the figure provides a detailed demonstration of how the toolkit can be utilized to analyze observational video data. Specifically, it illustrates an analysis generated by the toolkit from a video capturing a collaborative learning session. This visual representation highlights the toolkit's capabilities in processing and interpreting complex, real-time learning environments, thereby offering valuable insights into the dynamics of collaborative learning.

3. Discussion and Future Research Directions

The screenshot shows a web interface titled "Multimodal Generative Analysis". It features several input fields and a button. The "Upload Video" section has a "Choose File" button and a text box containing "1_minute_test.mov". The "Enter Prompt" section has a text box with a detailed prompt: "what are the students doing here? examine their interactions in the framework of multimodal interaction analysis. take the frames as whole, consider them a video and respond accordingly. give me a short answer." Below this are two input fields: "Frame sampling rate" with the value "140" and "Max Tokens" with the value "500". A blue button labeled "Send for Analysis" is positioned below these fields. At the bottom, a text area displays the analysis result, which is a paragraph describing student interactions based on video frames.

Figure 3: Generative Multimodal Analysis (GMA) Toolkit for Video Data

Generative Multimodal Analysis (GMA) represents a groundbreaking approach in the field of learning analytics, addressing the complexities of learning processes through the integration of generative AI and multimodal data. This approach is particularly relevant given the diverse nature of learning environments and the myriad forms of data they generate.

In learning analytics, the focus is traditionally on quantifiable data such as test scores, completion rates, and engagement metrics [9], [10]. However, the advent of GMA heralds a shift towards a more nuanced understanding of the learning process. By incorporating GenAI with MLLMs, GMA can seamlessly interpret and synthesize vast and varied datasets, including textual, auditory, and visual inputs, which are often overlooked in conventional analytics models. The inclusion of multimodal data is crucial for a comprehensive understanding of learning dynamics. As indicated by research in the field of educational technology, learning is not a unidimensional process but a complex interplay of cognitive, emotional, and social factors [11]. GMA's ability to analyze multimodal data allows for insights into these dimensions, offering a holistic view of the learning experience.

Future research endeavors should focus on conducting systematic examinations of the various components and procedural steps crucial for effective implementation of GMA. This direction of inquiry is essential to establish a set of clear, well-defined guidelines that can assist researchers in optimally employing GMA methodologies. Additionally, given the ethical complexities surrounding AI in education [12], there is a pressing need for comprehensive research aimed at establishing practical ethical guidelines for the application of GenAI methodologies in educational research. Such guidelines would not only streamline the application of GMA across diverse educational settings but also ensure that its integration into learning analytics is both efficient and impactful. By laying out these parameters,

future studies will significantly contribute to enhancing the utility and scalability of GMA, thereby advancing its role in educational technology, learning analytics, and learning sciences.

4. Acknowledgements

This research has been funded by the Research Council of Finland (aka. Academy of Finland) grants 350249, and the University of Oulu profiling project Profi7 Hybrid Intelligence - 352788.

5. REFERENCES

- [1] R. Luckin and M. Cukurova, 'Designing educational technologies in the age of AI: A learning sciences-driven approach', *British Journal of Educational Technology*, vol. 50, no. 6, pp. 2824–2838, Nov. 2019, doi: 10.1111/bjet.12861.
- [2] W. Holmes *et al.*, 'Ethics of AI in Education: Towards a Community-Wide Framework', *Int J Artif Intell Educ*, Apr. 2021, doi: 10.1007/s40593-021-00239-1.
- [3] S. Järvelä, A. Nguyen, and A. Hadwin, 'Human and artificial intelligence collaboration for socially shared regulation in learning', *British Journal of Educational Technology*, 2023.
- [4] H. Crompton and D. Burke, 'Artificial intelligence in higher education: the state of the field', *International Journal of Educational Technology in Higher Education*, vol. 20, no. 1, p. 22, Apr. 2023, doi: 10.1186/s41239-023-00392-8.
- [5] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, 'Systematic review of research on artificial intelligence applications in higher education – where are the educators?', *International Journal of Educational Technology in Higher Education*, vol. 16, no. 1, p. 39, Oct. 2019, doi: 10.1186/s41239-019-0171-0.
- [6] Y. K. Dwivedi *et al.*, "'So what if ChatGPT wrote it?' Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy', *International Journal of Information Management*, vol. 71, p. 102642, Aug. 2023, doi: 10.1016/j.ijinfomgt.2023.102642.
- [7] E. Kasneci *et al.*, 'ChatGPT for good? On opportunities and challenges of large language models for education', *Learning and Individual Differences*, vol. 103, p. 102274, Apr. 2023, doi: 10.1016/j.lindif.2023.102274.
- [8] S. Järvelä and M. Bannert, 'Temporal and adaptive processes of regulated learning - What can multimodal data tell?', *Learning and Instruction*, vol. 72, p. 101268, Apr. 2021, doi: 10.1016/j.learninstruc.2019.101268.
- [9] R. S. Baker, 'Challenges for the Future of Educational Data Mining: The Baker Learning Analytics Prizes', *Journal of Educational Data Mining*, vol. 11, no. 1, pp. 1–17, Jun. 2019, doi: 10.5281/ZENODO.3554745.
- [10] A. Nguyen, L. Gardner, and D. Sheridan, 'Data Analytics in Higher Education: An Integrated View', *Journal of Information Systems Education*, vol. 31, no. 1, pp. 61–71, Jan. 2020.
- [11] A. Nguyen, S. Järvelä, C. Rosé, H. Järvenoja, and J. Malmberg, 'Examining socially shared regulation and shared physiological arousal events with multimodal learning analytics', *British Journal of Educational Technology*, vol. 54, no. 1, pp. 293–312, 2023, doi: 10.1111/bjet.13280.
- [12] A. Nguyen, H. N. Ngo, Y. Hong, B. Dang, and B.-P. T. Nguyen, 'Ethical principles for artificial intelligence in education', *Educ Inf Technol*, vol. 28, no. 4, pp. 4221–4241, 2023, doi: 10.1007/s10639-022-11316-w.