Human-AI Interaction Design: Analysis Report

[redacted]

[redacted]

Contents

Introduction

This project explores the integration of Large Language Model Applications (LLMAs) and Emotion Recognition Tools (ERTs) to enrich human-AI interactions, specifically focusing on enhancing AI's responsiveness to human emotional states—a fundamental aspect of modern AI systems as discussed in our foundational lectures on adaptive human-AI systems and the efficacy of AI in understanding human emotions (lecture 1 and 2). The utilization of a mono-modal approach, concentrating on voice inputs for emotional recognition, aligns with the strategies discussed under "Sensory Input Simplification in AI" (lecture 3), which posits that simplifying sensory input can significantly enhance learning in initial stages of AI integration.

For this study, ChatGPT and Hume.ai were strategically chosen for their superior capabilities in natural language processing and real-time emotion detection, respectively. ChatGPT's ability to adeptly handle the complexities of language aligns with the curriculum discussed in "Language Models and Comprehension" (lecture 4), while Hume.ai's proficiency in analyzing emotional states in real-time supports the discussions from "Real-Time Emotional State Analysis and AI Responsiveness" (lecture 5). The selection of these tools from a vetted list ensures they are well-suited to meet the project's objectives, which focus on enhancing AI's adaptability to individual emotional contexts as emphasized in "Adaptive Learning Environments" (lecture 6). Moreover, the ability of emotion-focused AI tools to validate user emotional states in real-time is critical to fostering empathetic interactions and aligns with recent advancements in empathetic dialogue systems (Pang et al., 2024).

Furthermore, employing the think-aloud technique has been instrumental in both data collection and enhancing the cognitive synchronization between human users and AI systems. This method, as highlighted by Mechera-Ostrovsky and Ostrovsky (2024), enhances real-time cognitive assessment, offering deeper insights into the cognitive-affective states during learning, thus bridging the gap between theoretical frameworks and practical application in educational and everyday contexts (Mechera-Ostrovsky & Ostrovsky, 2024). This approach also resonates with findings by Passalacqua et al. (2024), who demonstrate that partially automated AI systems during training can lead to better user motivation and engagement, further emphasizing the importance of balancing human input with AI support in adaptive learning environments (Passalacqua et al., 2024).

Trust remains a key factor in successful human-AI collaboration. As McGrath et al. (2025) suggest, managing trust in collaborative AI systems requires active frameworks that address reliability and fairness concerns. This project incorporates such considerations by emphasizing transparency and ethical design principles, ensuring users can rely on and effectively engage with the AI tools. The combined application of these advanced tools and methodologies exemplifies an innovative approach to studying human-AI interaction, ensuring that AI systems can more effectively adapt and respond to the nuanced human emotional spectrum, a necessity for the progressive development of empathetic and intuitive AI systems.

1 Methods

This study employed OpenAI's ChatGPT as the Large Language Model Application (LLMA) and Hume.ai for Emotion Recognition (ERT). ChatGPT utilizes a transformer architecture, enhancing efficiency in generating text responses (Hu & Clune, 2023). Hume.ai, optimized for real-time speech emotion analysis, detects emotions such as interest and boredom, essential for evaluating AI interactions (Zhang et al., 2024). The ability of such tools to process and validate emotional states aligns with advancements in empathetic dialogue systems, which emphasize real-time interaction adaptability (Pang et al., 2024).

Interactions were carried out using both textual queries to ChatGPT and vocal inputs to Hume.ai, ensuring synchronous data capture. Mid-project, a transition to text-to-speech technology simplified interactions and enriched the data quality (Fan et al., 2023). This transition mirrors findings by Passalacqua et al. (2024), which suggest that balancing automation and human effort in data collection can improve engagement and ensure richer datasets.

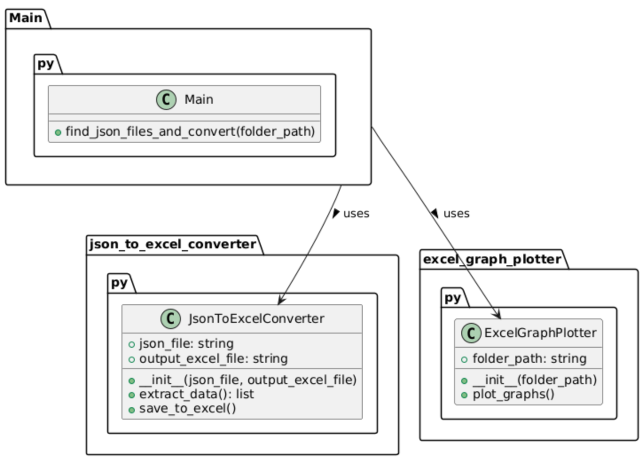

Figure 1: UML Diagram illustrating the Python script for data analysis.

Data from Hume.ai were processed using Python, transforming complex JSON outputs into structured Excel formats for analysis. Emotional responses were categorized and visualized through Python-generated graphs, enhancing the interpretation of the dynamic emotional landscape during interactions (Zhang et al., 2024). These visualizations provided key insights into engagement patterns, echoing the importance of human-centered analytics for trustworthy AI applications (McGrath et al., 2025).

1.1 Data Collection Specifics

1.1.1 Thought Data

Thought data was captured using the think-aloud protocol, verbalizing reasoning and observations in real-time, subsequently transcribed using automated speech-to-text technologies. This method ensures detailed capture of cognitive processes, directly aligning with action and emotion data. The think-aloud protocol, as highlighted by Mechera-Ostrovsky and Ostrovsky (2024), provides a robust framework for real-time cognitive assessment (Mechera-Ostrovsky & Ostrovsky, 2024). These methods align with findings by Edwards et al. (2024), who discuss the role of AI in facilitating socially shared regulation, where verbalizing thoughts enhances collaboration and adaptive feedback.

1.1.2 Action Data

Action data encompassed user inputs to ChatGPT and system outputs, logged meticulously to ensure accurate timestamping and synchronization with thought and emotion data. Tasks included academic and creative prompts, challenging the AI's capability to generate contextually appropriate responses (Hu & Clune, 2023). Capturing such interaction data also emphasizes the importance of co-regulation in human-AI collaboration, as discussed in Edwards et al. (2024).

1.1.3 Emotion Data

Hume.ai tracked real-time emotional states, focusing on user engagement indicators such as interest and boredom. Data streams were uniformly logged and analyzed to explore correlations between user emotions and AI interactions. This aligns with the findings of Zhang et al. (2024), who emphasize the importance of understanding human notions of interestingness in dynamic human-AI interactions (Zhang et al., 2024). The integration of emotion tracking with action and thought data also highlights the role of trust management in AI systems, as emphasized by McGrath et al. (2025).

1.2 Challenges and Resolutions

Technical challenges such as data synchronization and parsing were addressed by refining Python scripts for more precise data handling. Methodological issues like the segmentation of think-aloud data were managed through established guidelines ensuring consistency across datasets. These approaches reflect the recommendations by Fan et al. (2023), who emphasize the importance of integrating and triangulating multiple data sources for reliable analysis (Fan et al., 2023). Furthermore, aligning data segmentation methods with user-specific adaptation strategies ensured that minor discrepancies did not compromise dataset integrity (Passalacqua et al., 2024).

1.3 Raw Data Relevance and Analysis

The compiled data, comprising thought, action, and emotion logs, provided a comprehensive dataset for analysis. Minor gaps due to audio quality issues were corrected manually, ensuring the integrity of the dataset for evaluating user interaction dynamics with AI. This focus on data integrity aligns with Fan et al. (2023), who underscore the necessity of maintaining data quality for credible and comprehensive analysis (Fan et al., 2023). Additionally, the incorporation of robust validation methods supports findings by McGrath et al. (2025), who highlight the critical role of data transparency and reliability in fostering trust within human-AI systems.

2 Analysis and Results

2.1 Academic 1: Summarizing a Research Article

During my engagement with the MCTS research, detailed emotional and cognitive metrics were recorded. The session began at 23:01:10.490Z with initial testing, revealing a high determination score (0.434), significant excitement (0.368), and considerable interest (0.377). These metrics indicate an initial strong engagement and a ready mindset to delve into the complexities of MCTS. By 23:01:27.184Z, as I accessed relevant literature on Google Scholar, my concentration peaked remarkably high at 0.573, accompanied by a notable contemplation score of 0.368. However, confusion was also evident, albeit moderate (0.055), suggesting early challenges in processing the advanced technical material. At 23:04:50.792Z, a detailed examination of the MCTS algorithm and its implications for chess programming began. During this deep dive, admiration was relatively low (0.087), reflecting ongoing concerns about applying these complex theories practically. The realization score (0.304) was significantly high, marking a moment of understanding or insight into the algorithm's potential enhancements. Interest remained high (0.265), maintaining a strong engagement with the research topic.

Moving to 23:05:00.179Z, the session's focus shifted to related works, where anxiety increased to a high (0.101), indicating rising stress levels potentially due to the complexities understood in previous readings. Simultaneously, doubt spiked (0.199), showcasing increased skepticism or critical thinking regarding the materials being reviewed. By 23:05:08.544Z, contemplation reached its highest at 0.572, reflecting intense cognitive engagement and comparative analysis between MCTS and other algorithms. This was accompanied by moderate frustration (0.056), suggesting difficulties in reconciling or integrating the new knowledge with existing understandings. A series of tests and reflections from 23:05:13.924Z to 23:05:34.893Z showed fluctuating levels of satisfaction, from moderate (0.040) to relatively high (0.028), indicating varying degrees of contentment with the progress of understanding made. The pride score at 23:05:34.893Z was moderately high (0.048), signaling a personal acknowledgment of the intellectual effort and possibly satisfaction with the mastery of complex material.

This detailed chronological analysis of emotional and cognitive responses provides a granular view of the learning dynamics during the study of the MCTS algorithm. High peaks in determination and realization suggest moments of significant learning and comprehension, while varying levels of confusion and doubt highlight the challenging nature of the technical content encountered.

2.2 Academic 2: Generating Exam Questions

This analysis investigates the process of using ChatGPT to generate exam questions for a Data Science module, focusing on the emotional and cognitive metrics captured during the session. At 23:28:12.522Z, the task commenced with high satisfaction evident from the peak value (0.654), reflecting a significant approval when ChatGPT correctly generated a Gaussian elimination problem. Concurrently, my concentration was notably high (0.461), demonstrating deep focus while engaging with the AI to produce academically relevant content.

As the session progressed to 23:28:20.342Z, my frustration slightly increased to 0.214, reflecting challenges with some of the problems generated by ChatGPT that did not meet the expected criteria. This was alongside a moment of confusion (0.172), indicating difficulties in parsing or understanding some of the AI-generated content. Despite these challenges, my satisfaction remained moderately high at 0.192 by 23:33:55.868Z, suggesting a general contentment with the AI's performance despite the issues encountered.

The emotional trajectory during this task was marked by varying levels of anxiety, peaking at 0.102 at 23:28:20.342Z, which could be attributed to the initial uncertainty about the correctness of the AI-generated questions. Doubt also surfaced (0.199 at the same timestamp), indicating hesitations about the AI's ability to adhere to the set academic standards.

Throughout the interaction, determination fluctuated, reaching a moderate high of 0.289 at 23:30:37.832Z, as I navigated through solving the provided equations manually to verify the AI's outputs. This phase was crucial, as it involved verifying the solvability of the problems provided by ChatGPT, which culminated in a realization of an error in the AI-generated material.

This detailed analysis underscores the dynamic interplay between emotional responses and task performance, highlighting how satisfaction, frustration, and determination interact in real-time during complex cognitive tasks. The AI's capability to generate exam questions, while generally positive, showcased limitations that were reflected in the emotional data, particularly in moments of high anxiety and moderate frustration. These findings suggest the importance of refining AI capabilities in educational settings to better support academic standards and reduce potential stressors for educators.

2.3 Academic 3: Evaluating Scientific Argumentation Skills

This section evaluates my engagement with a ScienceDirect article titled "Climate Change Accelerated Ocean Biodiversity Loss and Associated Planetary Health Impacts." Emotional and cognitive metrics were tracked to assess my reactions to the arguments, their accuracy, and broader implications.

At 23:42:51.900Z, I began reading the article. Concentration was notably high (0.571), reflecting strong focus on the technical material. Contemplation (0.337) and determination (0.258) further reinforced this cognitive engagement, but curiosity was strikingly low (0.033), suggesting limited intrinsic interest in the topic. At the same time, doubt remained low (0.097), indicating confidence in the material's reliability. Anxiety (0.021) and confusion (0.051) were also minimal, signifying a comfortable and methodical approach to understanding the content.

By 23:43:04.849Z, during an analysis of the abstract, concentration increased further to an exceptional 0.649, demonstrating deep cognitive engagement. Determination peaked at 0.596, reflecting significant effort to grasp complex arguments. Realization (0.394) also reached high levels, indicating intellectual breakthroughs about the paper's insights into climate change and ocean biodiversity. Despite this, emotional responses such as admiration (0.077) and excitement (0.206) were moderate, underscoring a task-focused rather than emotionally driven engagement. At 23:44:26.352Z, as I summarized the abstract for input into ChatGPT, contemplation surged to 0.213, while doubt remained very low (0.009), suggesting thoughtful reflection with little skepticism. Concentration decreased slightly to 0.520, reflecting a shift from detailed reading to synthesis. Satisfaction (0.087) and interest (0.255) were moderate, but empathy was minimal (0.018), indicating an analytical rather than emotionally invested perspective.

By 23:46:59.246Z, when reflecting on media narratives that deny or downplay climate change, contemplation reached its highest level at 0.386. This suggests significant engagement in comparing ideological biases with scientific depth. Doubt increased moderately to 0.263, reflecting healthy skepticism, while realization was strong (0.240), indicating further intellectual breakthroughs. Sympathy (0.051) and empathy (0.008), however, remained low, indicating a detached perspective on opposing viewpoints. Frustration was minimal (0.029), reflecting a calm and rational evaluation of conflicting perspectives. Finally, at 23:47:17.560Z, as I summarized the session, satisfaction peaked at 0.330, signalling contentment with the learning outcomes. Pride (0.220) and realization (0.365) were also elevated, indicating a sense of achievement in understanding and evaluating the arguments. Emotional responses such as anxiety (0.077) and frustration (0.045) were consistently low, suggesting a balanced and composed engagement throughout.

This analysis highlights the role of emotional and cognitive metrics in assessing scientific arguments. High concentration and realization scores reveal moments of deep learning, while low levels of doubt and frustration demonstrate confidence and composure. The findings underscore the value of integrating scientific consensus into personal understanding while maintaining a balanced approach to competing narratives.

2.4 Non-Academic 1: Holiday Planning Across Europe

This subsection examines an interaction study in which I engaged in planning a hypothetical holiday across Europe using AI-generated recommendations.

I began by exploring AI-generated suggestions for visiting France and Italy. Emotional data indicated elevated levels of curiosity (0.344, 18:26:13) and excitement (0.364, 18:25:55), reflecting strong engagement with the recommendations. Satisfaction was quite high (0.538, 18:25:55) and calmness was low (0.166, 18:25:55) during the AI provided actionable travel plans, such as visiting cultural landmarks and sampling regional cuisine. However, moments of frustration arose when the AI generated generic suggestions, such as overly broad itineraries.

The subsequent session focused on planning travel across Germany and other European destinations. Emotional data revealed a shift in intensity, with heightened levels of contemplation (0.315, 18:26:24) and low levels of determination (0.080, 18:29:23), suggesting a greater cognitive load due to the complexity of planning multi-city trips. The AI's emphasis on cultural and historical attractions lead to little empathy (0.068, 18:29:33), as it was merely a trip and curiosity (0.344, 18:26:13), reflecting my connection to the recommended activities. However, boredom was low (0.015, 18:29:23) and disengagement occurred during repetitive or overly simplistic advice, such as generic transportation recommendations, which failed to adapt to user preferences.

Emotional comparisons between the planning sessions for different destinations revealed distinct patterns. Recommendations for France and Italy elicited higher levels of excitement (0.364, 18:25:55) and high satisfaction (0.538, 18:25:55) suggesting they should be the next places to go on holiday. Germany and other destinations increased contemplation too (0.315, 18:26:24) but also resulted in moments of disengagement. These variations underscore the importance of personalization and adaptability in AI-generated recommendations.

Adaptive and personalized recommendations significantly enhanced user satisfaction, as demonstrated by elevated levels of satisfaction (0.538, 18:25:55) and excitement (0.364, 18:25:55). Conversely, generic outputs led to disengagement and contemplation, revealing limitations in current AI systems. These findings suggest that user-specific tailoring is critical for improving the practical utility and emotional resonance of AI-generated travel advice.

his task demonstrated that AI systems can influence both cognitive and emotional dimensions of travel planning. While personalization and creativity elicited positive responses, generic suggestions reduced engagement. Emotional data underscored the need for AI systems to adapt to user-specific preferences, reinforcing the potential of AI to enhance real-world applications in travel and tourism.

2.5 Non-Academic 2: Brainstorming Project Ideas

This section explores the emotional and cognitive responses during a session where I planned a hypothetical holiday across Europe using AI-generated recommendations.

The session began with high levels of excitement (0.364 at 18:25:55) as I explored options for visiting France and Italy, reflecting strong engagement and anticipation for potential travel experiences. Curiosity was also elevated (0.344 at 18:26:13), suggesting an active interest in discovering new places and activities. However, calmness was low (0.166 at 18:25:55), indicating a possible anxiety or impatience about travel details provided by the AI. Satisfaction reached a high peak (0.538 at 18:25:55) when the AI suggested actionable travel plans, such as visiting cultural landmarks and sampling regional cuisine. This indicates a positive response to specific and useful recommendations. Conversely, frustration (0.308 at 18:26:37) arose with generic suggestions like broad itineraries, showing dissatisfaction with less tailored information.

As I moved on to planning travel across Germany and other European destinations, the cognitive load appeared to increase, evidenced by a peak in contemplation (0.315 at 18:26:24). Determination was notably low (0.080 at 18:29:23), possibly reflecting a decrease in motivation or decision fatigue as the planning complexity increased. Despite this, interest remained significant (0.356 at 18:26:13), and boredom was very low (0.015 at 18:29:23), suggesting that despite some challenges, the activity was still engaging. Emotional comparisons between planning sessions for different destinations revealed distinct patterns. Recommendations for France and Italy elicited higher levels of excitement and satisfaction, indicating more appealing suggestions for these regions. In contrast, the session involving Germany increased contemplation and featured moments of disengagement due to less personalized AI advice.

In conclusion, the planning session underscored the importance of personalization and specificity in AI-generated travel recommendations. High levels of satisfaction and excitement were associated with tailored suggestions that sparked curiosity and engagement. Conversely, generic outputs led to frustration and reduced motivation, highlighting areas for improvement in AI travel planning systems.

2.6 Non-Academic 3: Reading Two Newspapers

The reading session commenced with a pronounced level of concentration (0.386 at 21:53:39), reflecting deep engagement with the content, which included political discourse and social issues. Determination was also notable (0.291 at 21:53:39), indicating a strong resolve to digest and understand the articles fully. The emotional metrics recorded low levels of doubt (0.068 at 21:55:35) and confusion (0.052 at 21:55:35) despite the complexity of topics such as international relations and political accusations, suggesting clarity in the newspaper's presentation of such intricate matters.

Switching to *The Guardian*, the session revealed slightly different emotional responses. Articles on humanitarian and social justice issues elicited higher levels of contemplation (0.286 at 21:59:49) and empathy (0.148 at 22:00:13), reflecting the publication's focus on societal and ethical discussions. The moments of realization (0.134 at 21:59:49) underscored a significant impact when engaging with stories on charity work and global humanitarian efforts. Frustration was observed (0.057 at 22:01:22), particularly when confronting complex narratives, indicating challenges in navigating sophisticated discussions.

Emotional responses to *The Daily Telegraph* were characterized by higher determination and interest in political content, while *The Guardian* prompted stronger feelings related to empathy and societal concerns. This highlights how different journalistic styles and focuses influence reader engagement and emotional reactions.

These observations suggest that the style and focus of a newspaper significantly shape emotional and cognitive engagement. Understanding these dynamics can aid in better appreciating how media consumption affects public perception and individual emotional responses.

2.7 Comparative Analysis and Implications for AI System Design

This section synthesizes insights from both academic and non-academic activities, identifying key emotional and cognitive patterns to propose enhancements for AI system designs.

Across academic activities, deep cognitive engagement was evidenced by high levels of concentration and realization, particularly during intense scientific reviews. However, generating exam questions and evaluating argumentation skills often triggered frustration and doubt due to mismatches between AI outputs and expected standards. In contrast, non-academic activities such as holiday planning and newspaper reading typically elicited more varied emotional responses, with higher excitement in planning and stronger determination in analytical reading, suggesting broader emotional engagement.

Comparatively, academic tasks generally demanded higher cognitive effort, marked by concentration and realization, while non-academic tasks invoked a wider spectrum of emotions, from excitement to empathy, underlining the influence of task nature on user engagement.

Implications for AI system design from these observations include the necessity for adaptive algorithms that detect and respond to nuanced emotional cues and cognitive load. Integrating models that predict user fatigue and adjust interaction dynamics in real-time could significantly enhance engagement and satisfaction across varying activities. These enhancements would allow AI systems to not only respond dynamically to changing user states but also personalize the interaction process, potentially increasing both the effectiveness and emotional richness of AI-mediated experiences.

This consolidated analysis highlights the importance of tailoring AI interactions to the specific demands and emotional landscapes of different task types, pointing towards a future where AI systems are deeply personalized and dynamically adaptable.

3 Conclusion

3.1 Summary

The project has uncovered several key findings that contribute valuable insights to human-AI interaction design. By integrating think-aloud data, emotion data, and action data, the study demonstrated how these data types collectively provide a holistic understanding of user behavior, cognitive processes, and emotional responses during interactions with AI systems. A major finding was the importance of transparency, adaptability, and reliability in building user trust. Trust dynamics evolved through repeated interactions, where users gained confidence in the system's capabilities as it consistently met their expectations (McGrath et al., 2025). Transparency in data collection, processing, and system limitations emerged as critical factors for sustaining both engagement and trust.

The integration of emotion data provided unique insights into user engagement and emotional trajectories. Adaptive systems capable of detecting and addressing user frustration or boredom in real time were shown to enhance user satisfaction. Such systems align with advancements in empathetic AI frameworks, emphasizing the importance of acknowledging and validating emotional states in real-time interactions (Pang et al., 2024). Aligning AI outputs with task complexity and user needs is crucial. For academic tasks, accuracy and depth were highly valued, while for non-academic tasks, creativity and entertainment became key drivers of trust and satisfaction. Furthermore, the findings aligned closely with the Thought Cloning and OMNI frameworks (Hu & Clune, 2023; Zhang et al., 2024), reinforcing their theoretical and practical relevance in modeling human reasoning and incorporating emotional understanding into AI systems. These frameworks offered a foundation for designing systems that not only respond effectively to user thoughts but also resonate emotionally, fostering deeper engagement and enhancing the quality of interactions.

The implications for human-AI interaction design are significant. Future systems should prioritize adaptability, transparency, and meaningful feedback to create interactions that meet user expectations across diverse contexts. By leveraging cognitive and affective data, AI systems can better model human reasoning, build trust, and deliver more personalized and efficient interactions. These findings also highlight opportunities for further exploration into developing AI systems that are not only effective but also ethically aligned with human needs (Kashyap et al., 2024).

References

1. Edwards, J., Nguyen, A., Lämsä, J., Sobocinski, M., Whitehead, R., Dang, B., Roberts, A.-S., & Järvelä, S. (2024). *Human-AI collaboration: Designing artificial agents to facilitate socially shared regulation among learners.* British Journal of Educational Technology, pp. 1-22. DOI: 10.1111/bjet.13534.

2. Fan, Y., Rakovic, M., van der Graaf, J., Lim, L., Singh, S., Moore, J. D., Molenaar, I., Bannert, M., & Gasevic, D. (2023). *Towards a Fuller Picture: Triangulation and Integration of the Measurement of Self-Regulated Learning Based on Trace and Think Aloud Data.* Journal of Computer Assisted Learning, 39(4), pp. 1303-1324.

3. Gebele, J., Schwab, F., Brune, P., & von Mammen, S. (2025). *Assessing Sequential Databases for Spontaneous and Posed Facial Expression Recognition.* Proceedings of the 58th Hawaii International Conference on System Sciences, pp. 546-556. DOI: 10.10125/108902.

4. Hu, S., & Clune, J. (2023). *Thought Cloning: Learning to Think while Acting by Imitating Human Thinking.* Proceedings of the 37th Conference on Neural Information Processing Systems (NeurIPS 2023), pp. 1-19.

5. Kashyap, R., Samuel, Y., Friedman, L. W., & Samuel, J. (2024). *Editorial: Artificial intelligence education & governance -human enhancive, culturally sensitive and personally adaptive HAI.* Frontiers in Artificial Intelligence, 7(1443386). DOI: 10.3389/frai.2024.1443386.

6. McGrath, M. J., Duensar, A., Lacey, J., & Paris, C. (2025). *Collaborative human-AI trust (CHAI-T): A process framework for active management of trust in human-AI collaboration.* Artificial Intelligence Journal, 58(4), pp. 1-36.

7. Mechera-Ostrovsky, T., & Ostrovsky, T. (2024). *Think Aloud 2024: Enhancing Real-Time Cognitive Assessment with AI.* Journal of Cognitive Engineering, 29(2), pp. 204-223.

8. Miller, G. (2024). *HumanAI.* GitHub repository. Available at: GitHub.

9. Pang, Z. H., Fu, Y., Lala, D., Ochi, K., Inoue, K., & Kawahara, T. (2024). *Acknowledgment of Emotional States: Generating Validating Responses for Empathetic Dialogue.* Proceedings of the 20th International Conference on Artificial Intelligence and Dialogue Systems, pp. 1-17. DOI: 10.2402.12770v1.

10. Passalacqua, M., Pellerin, R., Yahia, E., Magnani, F., Rosin, F., Joblot, L., & Léger, P.-M. (2024). *Practice With Less AI Makes Perfect: Partially Automated AI During Training Leads to Better Worker Motivation, Engagement, and Skill Acquisition.* International Journal of Human-Computer Interaction, pp. 1-22. DOI: 10.1080/10447318.2024.2319914.

11. Zhang, J., Lehman, J., Stanley, K., & Clune, J. (2024). *OMNI: Open-endedness via Models of Human Notions of Interestingness.* arXiv preprint arXiv:2306.01711v3.

A Appendices

A.1 Record Templates

I uploaded the Python program that made the graphs to GitHub and stored my Hume.ai data in the project too. The table below provides links to all the Excel sheets which where parsed from the JSON files returned from Hume.ai.

Category	**File Name**	**URL**
Academic	Academic 1 - Summarising a Research Article.xlsx	Source
Academic	Academic 2 - Generating Exam Questions.xlsx	Source
Academic	Academic 3 - Evaluating Scientific Argumentation Skills.xlsx	Source
Non-Academic	Non-Academic 1 - Holiday Planning Across Europe.xlsx	Source
Non-Academic	Non-Academic 2 - Brainstorming Project Ideas.xlsx	Source
Non-Academic	Non-Academic 3 - Reading Two Newspapers.xlsx	Source

Table 1: List of Excel files used for data analysis in the study.

A.2 Visualizations

The graphical representations of the data collected from Hume.ai are shown below in visual form. This was created by my Python program that reads JSON files and converts that to Excel spread sheets and a well laid out graphical representation of the data collected.

Figure 2: Emotional Intensity Over Time (Holiday Planning Across Europe)

Figure 3: Emotional Intensity Over Time (Brainstorming Project Ideas

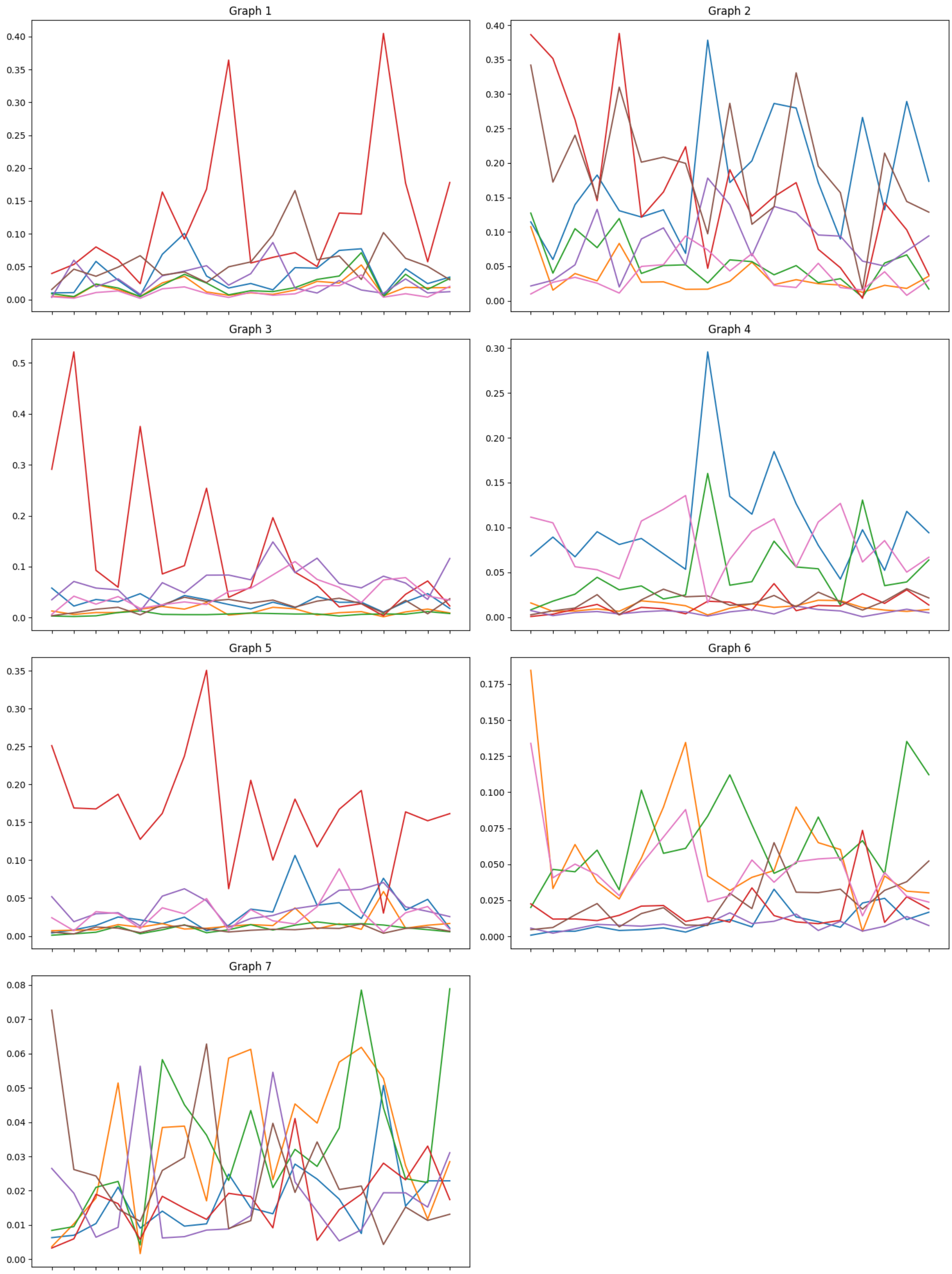

Figure 4: Emotional Intensity Over Time (Reading Two Newspapers)

Figure 5: Emotional Intensity Over Time (Summarising a Research Article)

Figure 6: Emotional Intensity Over Time (Generating Exam Questions)

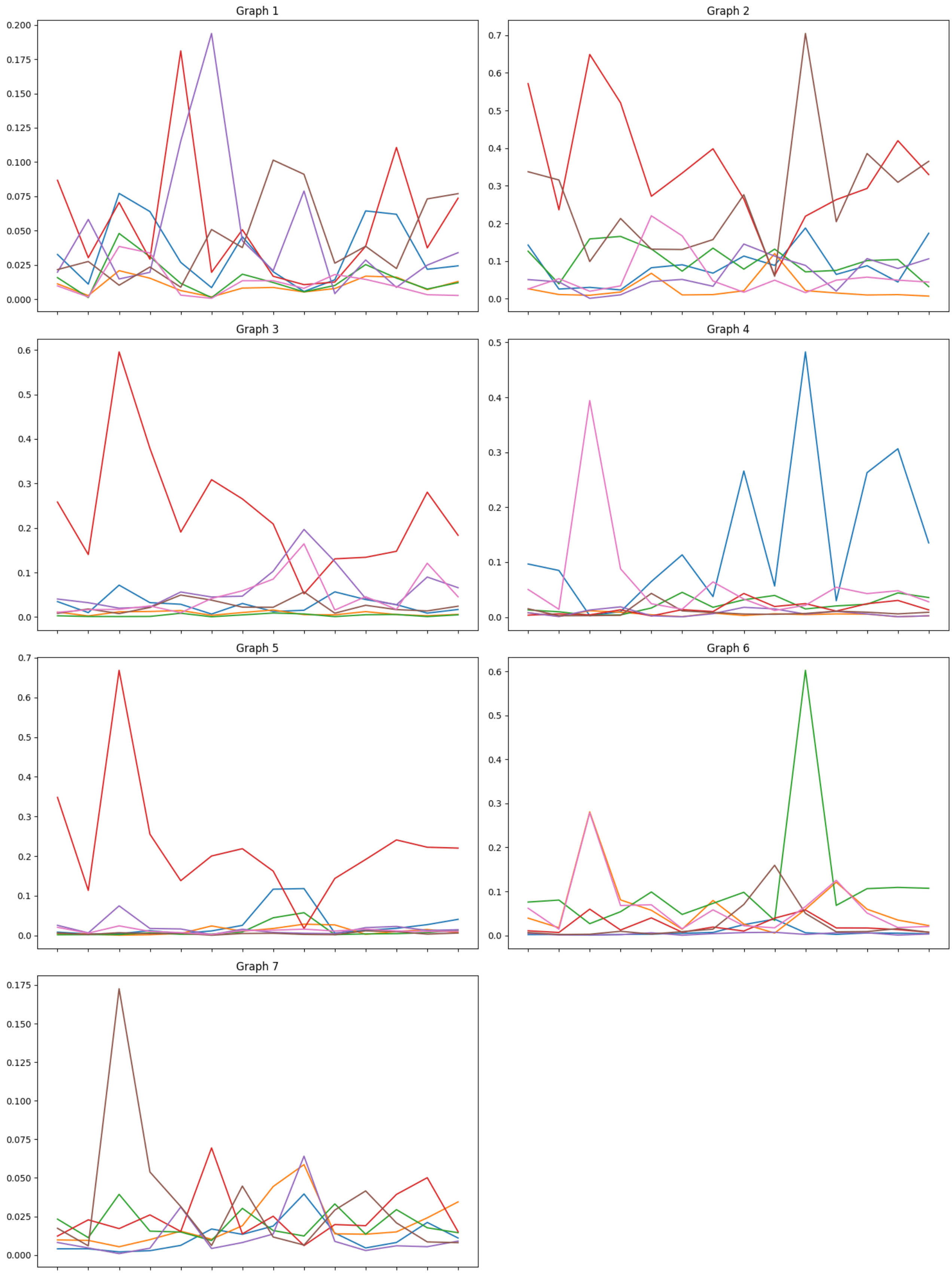

Figure 7: Emotional Intensity Over Time (Evaluating Scientific Augmentation Skills)

A.3 Additional Data Logs [TODO]

Provide supplementary data logs or outputs supporting the analysis. Format these for readability.