Iowa Initiative for Artificial Intelligence
Final Report

Project title:	Using Artificial Intelligence and Machine Learning (AI/ML) to Analyze Discourse from the Adult Attachment Interview.
Principal Investigator:	Lane Strathearn MD PhD
Prepared by (IIAI):	Avinash Reddy Mudireddy
Other investigators:	Jacob Michaelson PhD
Date:	July 15, 2025

Were specific aims fulfilled:	Partially
Readiness for extramural proposal?	Yes
If yes … Planned submission date	
Funding agency	
Grant mechanism	
If no … Why not? What went wrong?	

Brief summary of accomplished results:

This project successfully demonstrated a viable pathway for using advanced AI to analyze complex clinical interviews for adult attachment. While initial goals of direct classification were not met with older models, the project pivoted to a more sophisticated, multi-stage Retrieval-Augmented Generation (RAG) pipeline. This system proved highly effective at the foundational task of processing the interview transcripts. Specifically, it demonstrated a 98% accuracy in correctly identifying and grouping the question-and-answer (QA) pairs within the dialogue and an 80% accuracy in correctly categorizing those pairs according to the research template. These precise results validate the use of Large Language Models for structured information extraction from AAI transcripts and establish a robust framework for the future goal of full attachment classification.

Research report:

Aims (provided by PI):

- To explore the capabilities of AI/ML in coding complex, transcribed life narratives from the Adult Attachment Interview (AAI).
- To utilize this AI/ML model to predict adult attachment patterns in parents and examine their relationship with other biological data, such as brain imaging and hormonal responses.
- To foster a long-term, interdisciplinary collaboration between clinical researchers and machine learning experts.

The project's aims underwent a significant strategic evolution. The initial goal was to predict adult attachment patterns directly from interview transcripts. However, early experiments with models like BERT(Roberta model) yielded unsatisfactory performance (AUROC: 0.61). This led to a re-evaluation in consultation with the PI. As LLMs became more powerful, the team decided to break down the complex classification goal into smaller, more manageable tasks. The revised, sequential approach was:

1. First, perfectly segregate the Question and Answer (QA) pairs between the interviewer and interviewee.
2. Second, for each QA pair, identify specific "discourse markers" in the interviewee's response.
3. Third, use these identified markers to classify the adult attachment pattern. This change in strategy was driven by the data and the rapid advancements in AI, allowing for a more robust and ultimately more promising path toward the original goal.

Data:

The research was based on a rich and unique dataset of over 103 Adult Attachment Interviews collected over two decades by Dr. Strathearn. These interviews were conducted with women during pregnancy or in the postpartum period. Each interview, lasting 1.5-2 hours, was transcribed (yielding 15-20 pages of text) and manually coded by a team of reliable coders at the Family Relations Institute. This dataset is also linked to extensive biobehavioral data, including fMRI scans of maternal brain responses to infant cues.

AI/ML Approach:

The project explored three distinct AI/ML methodologies in chronological order:

1. **BERT-Based Direct Classification(**roberta_model)**:** The initial approach utilized a pre-trained BERT (Bidirectional Encoder Representations from Transformers) model. The entire interview transcript was input into the model to directly predict one of the adult attachment classifications. This method was an attempt at end-to-end classification without intermediate steps.
2. **Doc2Vec with GRU-Attention Network:** The second approach involved a more nuanced representation of the text. Each interview was converted into a numerical vector using the Doc2Vec algorithm. These document embeddings were then fed into a Gated Recurrent Unit (GRU) neural network, enhanced with an attention mechanism. The attention layer was designed to help the model weigh the importance of different parts of the interview when making a prediction.
3. **Retrieval-Augmented Generation (RAG) Pipeline:** The final and most successful method was a sophisticated, multi-step RAG pipeline built on the **Microsoft Azure** cloud platform. This system was designed for both security (HIPAA-compliance) and performance.
    - **Architecture:** The pipeline orchestrated several Azure services.

      Azure Storage was used for secure, encrypted storage of the interview transcripts.

      Azure OpenAI provided access to state-of-the-art LLMs like **GPT-4o** and **O4-mini**. The entire workflow was managed using **Azure PromptFlow**, which allowed for the chaining of different steps and the integration of prompts with the LLMs.

    - **Process:** The pipeline's first implemented task was the segregation of QA pairs. This involved feeding the transcript to the LLM with a specific prompt instructing it to identify and separate the interviewer's questions from the interviewee's answers.
        1. **Conversation Structuring:** First, the raw interview text from the .docx files was automatically parsed. The system identified the speaker for each paragraph (Interviewer or Subject) and grouped consecutive paragraphs from the same

speaker into a single conversational turn. This converted the unstructured document into a structured list of who said what, and in what order.

2. **Dynamic Prompt Construction:** Next, the entire structured conversation was embedded within a carefully engineered master prompt. This prompt also included a complete, JSON structure based on the official AAI template. The system was instructed to act as a clinical expert and fill this template.
3. **LLM-Powered Data Extraction:** The complete prompt, containing both the conversation and the template, was sent to an LLM (GPT-4o or O4-mini) via the **Azure OpenAI service**. The LLM's task was to read the entire dialogue and accurately extract the interviewee's answers, placing them into the corresponding fields of the JSON template.
4. **Structured Output Generation:** The pipeline's final output was the JSON object returned by the LLM, now fully populated with the extracted interview data. This file provided a structured, machine-readable representation of the interview content, ready for quantitative analysis.

Experimental methods, validation approach:

All experiments were performed on a dataset of **99 transcribed and coded Adult Attachment Interviews (AAIs)**. The methodologies evolved over the course of the project, each with its own specific architecture and validation approach.

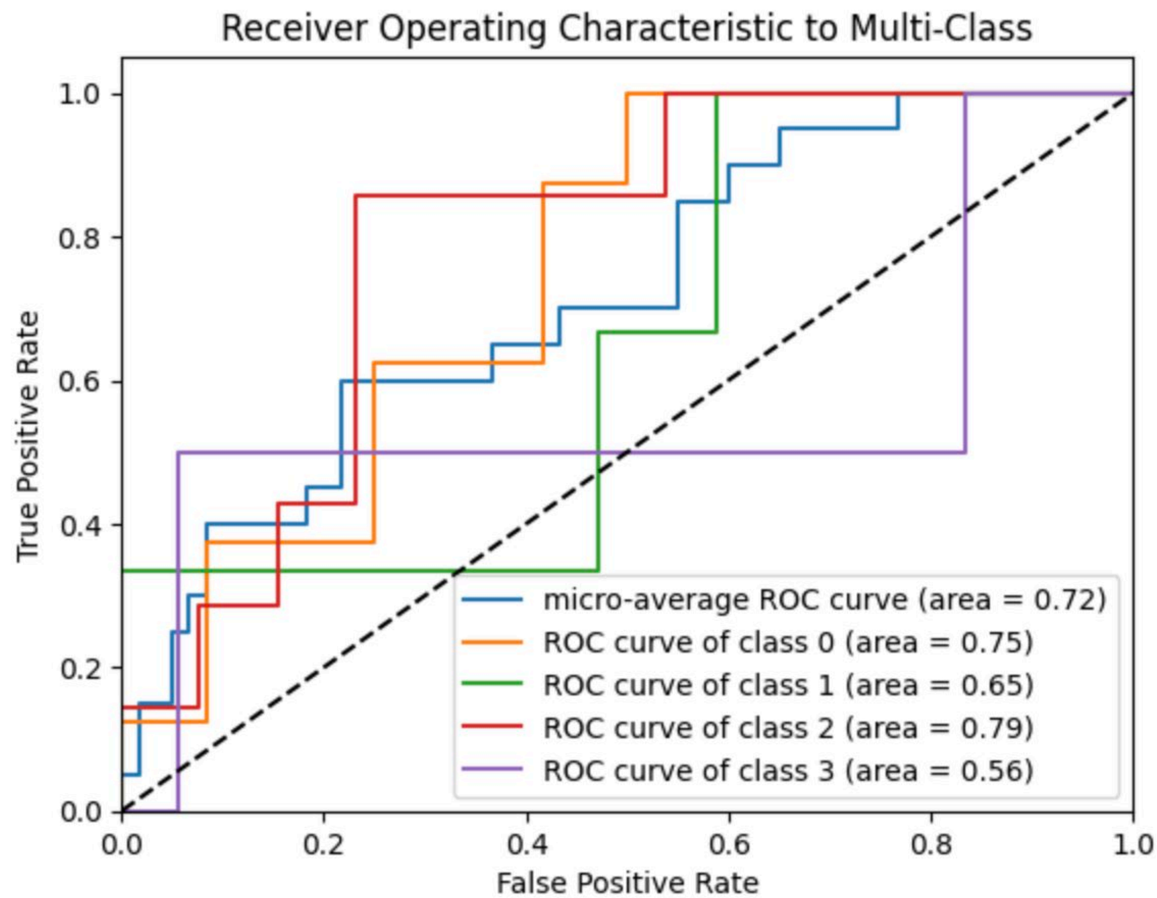
- **Stage 1: Direct Classification with BERT(**roberta_model)
  - **Network Structure:** A pre-trained **BERT (Bidirectional Encoder Representations from Transformers)**model, specifically the bert-base-uncased variant, was used for this initial phase. The model was set up for sequence classification with a single linear output layer to predict one of the AAI categories. The maximum sequence length for the input transcripts was capped at 512 tokens.
  - **Validation Approach:** The dataset was split into a **training set (80%) and a testing set (20%)**. The model was fine-tuned on the training set to learn the patterns associated with different attachment styles. Its performance was then evaluated on the held-out testing set to assess its ability to generalize to unseen data.
- **Stage 2: Document Embeddings with a GRU-Attention Network**
  - **Network Structure:** This approach first converted each transcript into a fixed-size vector using a pre-trained **Doc2Vec** model. These document embeddings were then fed into a custom-built neural network consisting of:
    - A **Gated Recurrent Unit (GRU)** layer to capture sequential patterns and context within the narrative.
    - An **Attention mechanism** built on top of the GRU. This layer allowed the model to dynamically weigh the importance of different parts of the interview, focusing on the most salient sections when making a classification decision.
    - A final **fully connected (linear) layer** that produced the final classification.
  - **Validation Approach:** Similar to the BERT stage, this model was trained and evaluated using an **80:20 train-test split** of the dataset to ensure a fair and robust assessment of its performance.
- **Stage 3: Retrieval-Augmented Generation (RAG) Pipeline**
  - **Network Structure:** This stage did not involve model fine-tuning. Instead, it leveraged the zero-shot capabilities of large language models (LLMs) through a carefully designed pipeline on **Microsoft Azure**. The core of this system was a series of structured prompts

sent to the O**4-mini** models via the Azure OpenAI service. The process was orchestrated using Azure PromptFlow, which managed the flow of data from Azure Storage, through the LLM, and to the final output. The "network" in this case is the orchestrated workflow of API calls and prompt templates designed to break down the analysis into sequential steps, starting with QA segregation.

- **Validation Approach:** Since no fine-tuning was performed, the concept of a train-test split did not apply. Instead, the entire dataset of 103(Expanded a few) interviews was used for evaluation. The performance of the pipeline's QA segregation task was measured by comparing the LLM's output directly against a manually verified "ground truth" version of the interviews. The key metrics calculated were:
    - **avg_overall_similarity_B**: Measured the ability of the model to correctly group a question with its corresponding answer.
    - **avg_sim_common_keys_A_prime**: Measured the ability of the model to correctly place those identified QA pairs into the appropriate categories as defined by the research template.

Results:

1) BERT-Based Direct Classification(roberta_model):
The initial BERT model for direct classification achieved a low AUROC of 0.61, indicating performance barely better than random chance.
2) Doc2Vec with GRU-Attention Network



	precision	recall	f1-score	support
0	0.60	0.38	0.46	8
1	0.33	0.33	0.33	3
2	0.50	0.43	0.46	7
3	0.17	0.50	0.25	2
accuracy			0.40	20
macro avg	0.40	0.41	0.38	20
weighted avg	0.48	0.40	0.42	20

model showed a notable improvement, with the AUROC increasing to 0.72. However, the precision and recall metrics were not sufficient for reliable classification. Class 0,1,2,3 here represents A,B,C,A/C

3) Retrieval-Augmented Generation (RAG) Pipeline:
focused on the QA segregation task, produced highly promising and nuanced results:

- It achieved **97.95% accuracy** (avg_overall_similarity_B) in the ability to correctly identify that a question and its corresponding answer belong together as a pair.
- It achieved **80.28% accuracy** (avg_sim_common_keys_A_prime) in the more difficult task of classifying those QA pairs into their correct respective bins according to the analysis template. These granular results demonstrate a strong capability for extracting and correctly structuring the dialogue, validating the pipeline's architecture as a solid foundation for the subsequent steps of discourse marker identification and final classification.

Note that several experiments were conducted throughout the project, but the major changes described above are the most significant.

Ideas/aims for future extramural project:

The success of the RAG pipeline provides a clear path for future work. The next phase will focus on completing the remaining steps of the analysis pipeline:

1. **Develop and Refine Prompts for Discourse Marker Identification:** Work with clinical experts to create a detailed, operationalized list of discourse markers. Then, engineer and test prompts for the LLM to accurately identify these markers in the interview text.
2. **Implement the Final Classification Layer:** Once the discourse markers are reliably extracted, develop a final model (which could be another LLM prompt or a more traditional classifier) that uses the presence and pattern of these markers to assign an overall attachment classification.
3. **Seek Extramural Funding**