Mining WhatsApp Course Groups for Sentiment Patterns

Team Member Information

Natanel Richey, natanel.richey@mail.huji.ac.il, ID: 332265842, CSE: nr33226584

Yehonatan Ezra, yonatan.ezra@mail.huji.ac.il, ID: 206762163, CSE: luna

Jonatan Vider, jonatan.vider@mail.huji.ac.il , ID: 204636286, CSE: viderspace

Problem Description

Motivation - As a student, especially at the start of the degree, whatsapp groups are a common way to get information and assistance while building a network within a large student population. These social network platforms can be extremely powerful, however many students do not effectively navigate them and receive responses they desire. We aim to explore two aspects of large group data.

In the first aspect, relating to sentiment analysis, we hypothesize that different university departments consist of different student personalities and consequently, a unique style of usage. Specifically, we hypothesize that the usage in Psychology groups is superior, leading to better engagement, collaboration and support framework. In contrast, usage amongst CS students is an example of how the platform should not be used.

In the second aspect, we hypothesize a clear and trending decline in average student mood over the time frame of an entire semester (weekly). The rationale behind this is our personal experience as students experiencing gradual stress (resulting from increasing academic workload) that peaks towards the end of the semester/beginning of the exam period. From this, we hypothesize that we will also recognize an expression of this phenomenon in the communication channels of students in their course groups. A gradual decline in general mood, an increase in stress levels, and even fluctuations in students' willingness to help others and express support/gratitude.

Data

Description:

We used WhatsApp group chat exports from various academic and social groups at Hebrew University and related communities. The dataset consists of 28 different WhatsApp groups organized into three categories: Computer Science courses (14 groups), General courses and university-related groups (6 groups), and Psychology & Biology courses (8 groups). Each group chat was exported as a structured JSON file containing complete conversation histories with metadata including group names and id, participant count, message count, and reactions. Additionally, rich message data was collected including message id, timestamp, body, replyTo and emoji reactions.

Implementation:

We developed a custom WhatsApp crawler application that goes beyond WhatsApp's basic export functionality.

The system uses the @open-wa/wa-automate API library (originally designed for chatbots) to access WhatsApp's internal data structures and extract comprehensive group information. Our custom JavaScript application includes several advanced features:

- Raw Data Extraction: The crawler accesses WhatsApp's API to extract raw JSON data containing messages, participants, and metadata not available through standard exports.
- Data Deduplication and Identity Resolution: The raw data contained inconsistencies—same group members appeared with different identifiers across messages. Our enrichment system (enrichment.js, participantsEnricher.js) automatically identifies and merges duplicate participants, resolving conflicts between phone numbers, user IDs (LIDs), and display names.
- Reply Thread Detection: Unlike basic exports that lose reply relationships, our system detects and links explicit reply messages to their original messages using WhatsApp's internal quotedMsg and quotedStanzaID references.
- Reaction Mapping: The crawler extracts and formats emoji reactions with full metadata, including reaction counts and participant identifiers.
- Stable Build for Distribution: The final version of the crawler was packaged as a stable build to ensure it could be easily distributed and run by students across different environments without requiring deep technical setup.

<u>Data Size:</u>

- Number of records: 23,619 total messages across all groups
- Number of groups: 28 WhatsApp groups
- Total participants: 7,061 participants (from metadata)
- File size: 12.78 MB total across all JSON files

Solution

As a first crucial step to testing our hypotheses, we performed a sentiment analysis on the whatsapp export data. We built a pipeline that exported the chat data to the GPT-4o-mini Language Model, which performed a sentiment analysis on a batch of 3 messages and exported the output as a JSON file. The LLM provides comprehensive annotations across multiple sentiment dimensions.

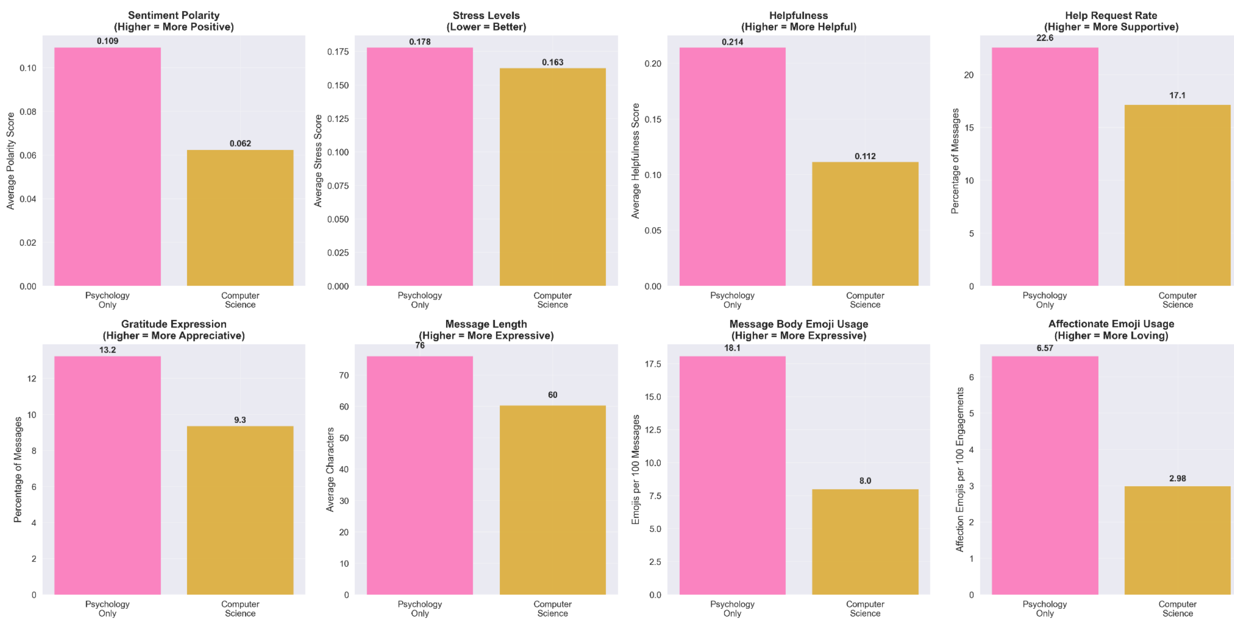
Prompt Engineering: *"You are a precise annotator for academic WhatsApp group chats (Hebrew & English). Return a JSON ARRAY where each element corresponds to ONE input message in the same order. Each element must be an object with fields: polarity: float in [-1.0, +1.0] (overall pleasantness), emotion_primary: one of {stress, gratitude, confusion, neutral_info, humor, anger, excitement, other}, emotion_summary: 1–2 words free-form (e.g., "stressed", "thankful", "confused", "calm", "excited"), stress_score: float [0..1] (urgency/pressure), uncertainty_score: float [0..1] (doubt/confusion), help_request: boolean (explicit ask for help), helpfulness: float [0..1] (contribution to solving), gratitude: boolean (thanks or ❤️/🙏), toxicity_score: float [0..1] (hostile language), info_drop: boolean (links/dates/official notices), reaction_sentiment: leave as null; caller may fill, evidence_terms: up to 5 short spans copied from the message (e.g., "שאלה", "❤️", "תודה 5", "http", "???", "דחוף", "link")."*

In support of our first hypothesis, we demonstrate several claims conducive to our narrative:

1. Positivity - positive and upbeat message sentiment.
2. Stress Management - lower stress levels and better coping mechanisms
3. Collaborative Support - more helpfulness and help-seeking behavior
4. Peer Recognition - higher gratitude expression and appreciation
5. Communication Quality - more verbose, expressive messaging
6. Emotional Intelligence - warmer, more supportive communication patterns through extensive emoji usage

In our script written to test these aspects we calculated the effect size with the def calculate_effect_size(group1_values, group2_values) function and the percentage difference with the def calculate_percentage_advantage(psych_mean, cs_mean) function.

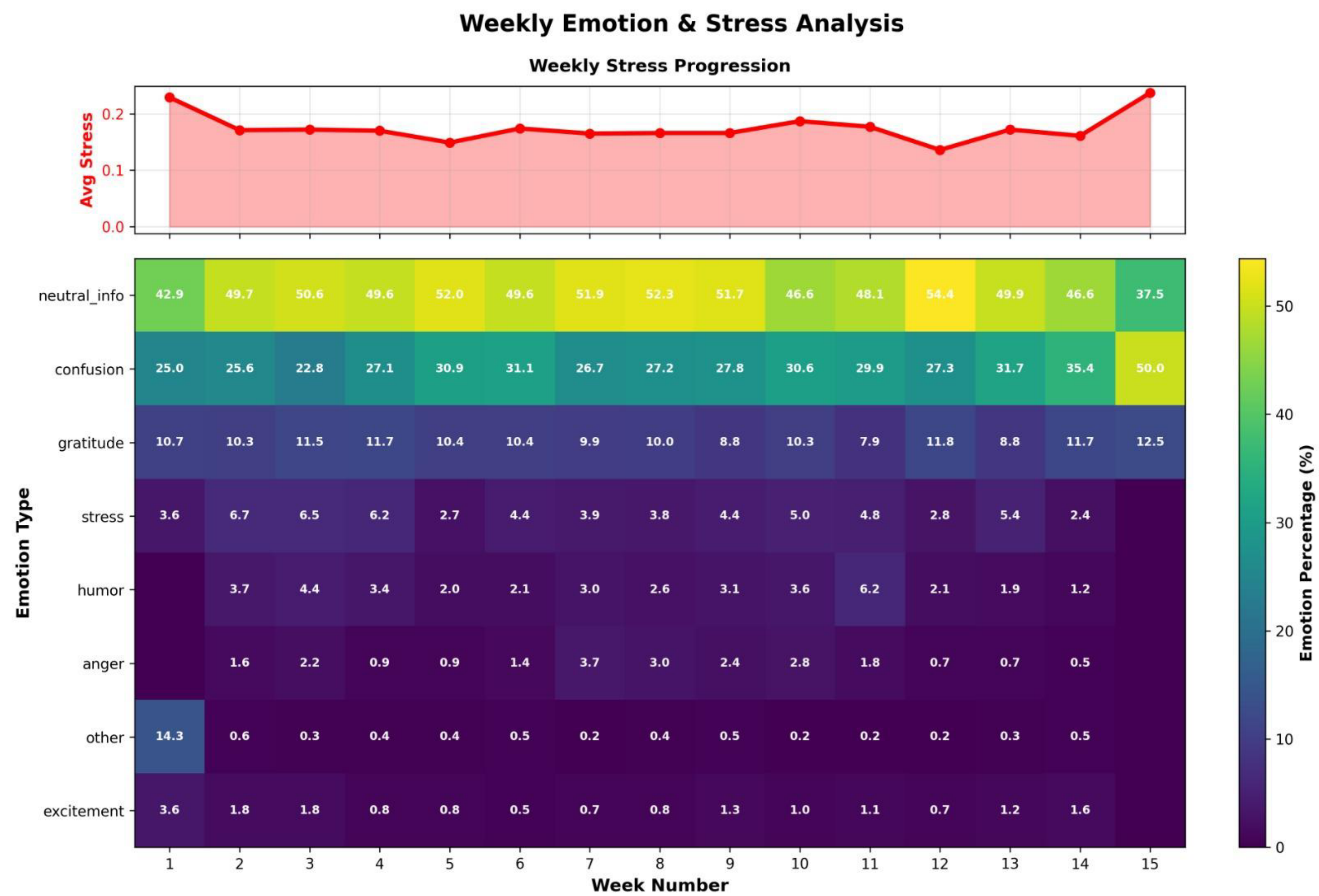
The results are best displayed by the following graphs:



Additionally, the following table clearly shows the analysis results:

Metric	Psychology Value	CS Value	Psychology Advantage (%)	Direction
Polarity (Positive Sentiment)	0.109291	0.062292	75.45	Higher is better
Stress Levels	0.178022	0.162601	9.48	Lower is better
Helpfulness	0.214336	0.111517	92.2	Higher is better
Help Request Rate	0.225774	0.171247	31.84	Higher is better

Gratitude Expression	0.132368	0.093483	41.59	Higher is better
Message Length (Characters)	76.00599	60.26518	26.11	Higher is better
Message Body Emojis	18.07	7.99	126.15	Higher is better
Affectionate Emojis	6.57	2.98	120.46	Higher is better



We visualized weekly sentiment patterns from WhatsApp course-group data to evaluate the hypothesis of declining student mood over the semester. A heatmap of emotion distributions across 15 weeks, combined with a stress timeline, revealed that all emotions didn't present a trend of any kind across the span of the semester. Contrary to expectations of decreasing positivity and rising stress near semester end, the data showed stable emotional patterns with no pronounced trends.

Evaluation

From the outset, we recognized a core challenge in this task - evaluating the quality of an assigned “emotion” label is not straightforward, whether that label comes from a human annotator or from an LLM. To address this, we set out to design practical, common-sense methods that could help us gauge how reliable the AI-generated sentiment labels were for each message.

Evaluation Criteria: We defined success as obtaining sentiment annotations that matched human intuition at the message level, remained consistent across different AI-models and aligned with real-world events such as exam periods in the academic calendar.

Setup: First, we scanned the raw JSON annotations to check that the labels made sense.
Second, we manually labeled a subset of messages and compared them with the model outputs.
Third, we cross-checked the results of two models, GPT-4o-mini and GPT-5, to see whether they produced consistent patterns. Finally, we aligned the temporal patterns of polarity with the academic calendar.

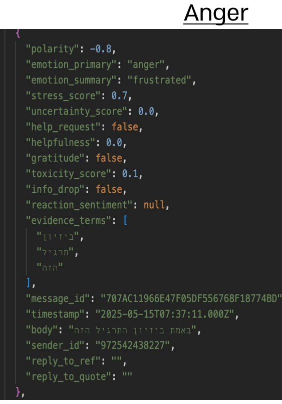
Result: The sanity check confirmed that messages were generally assigned reasonable polarity and primary emotions. Manual comparisons showed good consistency between our labels and the models, with most disagreements limited to ambiguous or mixed-emotion cases. Cross-model comparison revealed similar polarity patterns for major emotions with only small shifts. Temporal analysis showed polarity drops during exam periods (July-August) and recovery to moderately positive levels afterwards, reflecting expected academic stress cycles.

Visualization 1 - Illustrating Sanity Check Results:

High Polarity

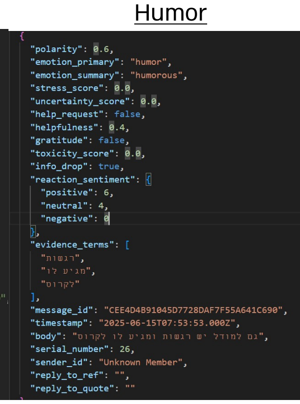
```
{
  "polarity": 1.0,
  "emotion_primary": "gratitude",
  "emotion_summary": "thankful",
  "stress_score": 0.0,
  "uncertainty_score": 0.0,
  "help_request": false,
  "helpfulness": 0.0,
  "gratitude": true,
  "toxicity_score": 0.0,
  "info_drop": false,
  "reaction_sentiment": {
    "positive": 0,
    "neutral": 1,
    "negative": 0
  },
  "evidence_terms": [
    "וואו",
    "אלופים",
    "תותחים",
    "❤️"
  ],
  "message_id": "A8CF4F2E4616D389834EF6964D046E1C",
  "timestamp": "2025-06-18T10:05:35.000Z",
  "body": "וואו אתם אלופים אגודה של תותחים אין עליכם ❤️",
  "serial_number": 322,
  "sender_id": "972543165655",
  "reply_to_ref": "",
  "reply_to_quote": ""
},
```


Anger



```
{
  "polarity": -0.8,
  "emotion_primary": "anger",
  "emotion_summary": "frustrated",
  "stress_score": 0.7,
  "uncertainty_score": 0.0,
  "help_request": false,
  "helpfulness": 0.0,
  "gratitude": false,
  "toxicity_score": 0.1,
  "info_drop": false,
  "reaction_sentiment": null,
  "evidence_terms": [
    "ביזיון",
    "תרגיל",
    "הזה"
  ],
  "message_id": "707AC11966E47F05DF556768F18774BD",
  "timestamp": "2025-05-15T07:37:11.000Z",
  "body": "באמת ביזיון התרגיל הזה",
  "sender_id": "972542438227",
  "reply_to_ref": "",
  "reply_to_quote": ""
},
```


Humor

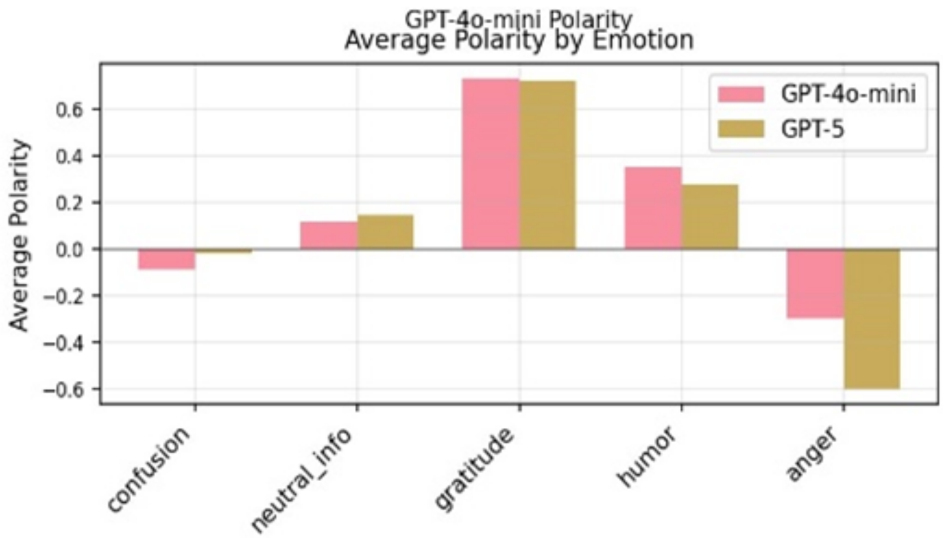


```
{
  "polarity": 0.6,
  "emotion_primary": "humor",
  "emotion_summary": "humorous",
  "stress_score": 0.0,
  "uncertainty_score": 0.0,
  "help_request": false,
  "helpfulness": 0.4,
  "gratitude": false,
  "toxicity_score": 0.0,
  "info_drop": true,
  "reaction_sentiment": {
    "positive": 6,
    "neutral": 4,
    "negative": 0
  },
  "evidence_terms": [
    "רגשות",
    "מגיע לו",
    "לקרוס"
  ],
  "message_id": "CEE4D4B91045D7728DAF7F55A641C690",
  "timestamp": "2025-06-15T07:53:53.000Z",
  "body": "גם למודל יש רגשות ומגיע לו לקרוס",
  "serial_number": 26,
  "sender_id": "Unknown Member",
  "reply_to_ref": "",
  "reply_to_quote": ""
```

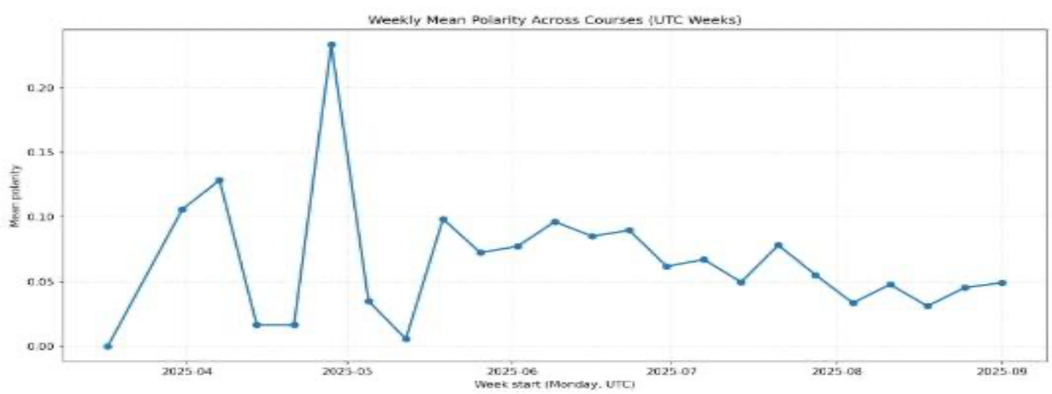

Gratitude

```
{
  "polarity": 0.9,
  "emotion_primary": "gratitude",
  "emotion_summary": "thankful",
  "stress_score": 0.0,
  "uncertainty_score": 0.0,
  "help_request": false,
  "helpfulness": 0.0,
  "gratitude": true,
  "toxicity_score": 0.0,
  "info_drop": false,
  "reaction_sentiment": {
    "positive": 3,
    "neutral": 2,
    "negative": 0
  },
  "evidence_terms": [
    "העוגן",
    "היציב",
    "בחיים",
    "של כולנו",
    "כרגע"
  ],
  "message_id": "3A8FBF275FFBA00168FD",
  "timestamp": "2025-06-15T00:08:26.000Z",
  "body": "אתה העוגן היציב היחיד בחיים של כולנו כרגע!",
  "serial_number": 19,
  "sender_id": "972587897849",
  "reply_to_ref": "",
  "reply_to_quote": ""
},
```

Visualization 2 - Model Comparison Plots:



Visualization 3 - Polarity Over Time



Future Work

We propose to use the tool we developed (we actually plan on doing this) on large public whatsapp groups for Junior programming positions. We will analyse the frequency of posts across different timespans (hour of day, day of week, month in year, etc), the description and requirement of the positions, recurring positions and other available information. Our goal is to gain a deeper understanding of Junior positions and potentially assist CS graduates in finding jobs.

Brief Conclusion

In this project, we analyzed WhatsApp course groups to understand student interaction patterns. Using a well-designed custom crawler application and sentiment analysis on over 23,000 messages. Evaluation through manual checks, subset labeling, and alignment with academic timelines confirmed that the sentiment annotations were generally reliable. We found strong support for the first hypothesis: psychology groups showed more positivity, helpfulness, and emotional engagement than computer science groups. However, the second hypothesis—that student mood declines across the semester—was not supported, as no clear temporal effect was observed.