# Med-Mingle: AI-Enhanced Content Moderation & Peer Support in Digital Health Communities

**Linda Berhe**
Advised by Heather Guarnera

## Project Overview

**Med-Mingle addresses critical gaps in digital health platforms through AI-enhanced content moderation and personalized peer matching.**

### Problem:

Digital health platforms struggle with unsafe content and disconnected user experiences.

### Solution:

BERT-based content moderation and hybrid recommendation algorithms create safer, more personalized health communities.

## Research Approach

### Motivating Question

How can AI enhance moderation and peer matching in health communities?

### Methodology

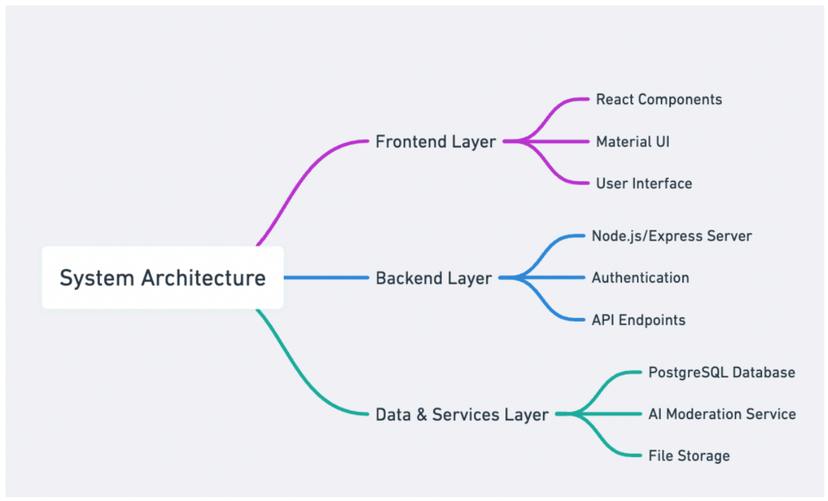
Integration of BERT-based NLP and Multi-Armed Bandit algorithms

### Key Findings

92% moderation accuracy

85% peer matching quality score

## System Architecture




## Content Moderation System

### BERT Model Implementation

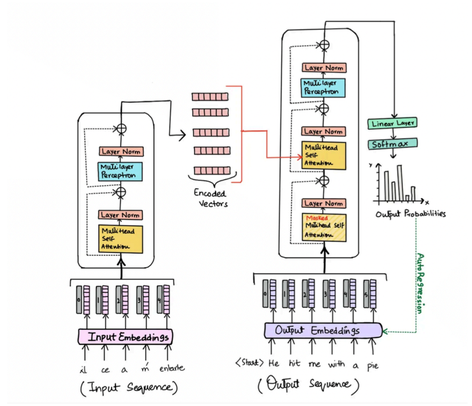
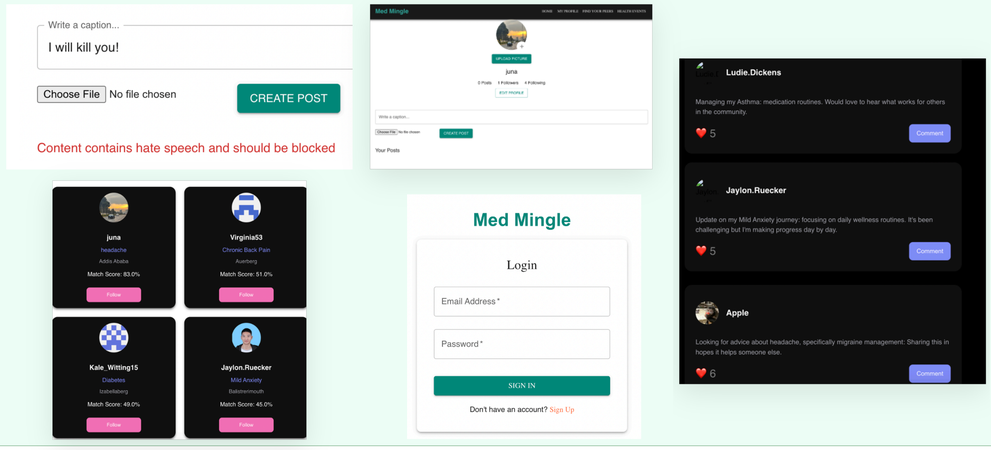



Figure 2: Self-Attention Mechanism in Transformers

### User Interface design



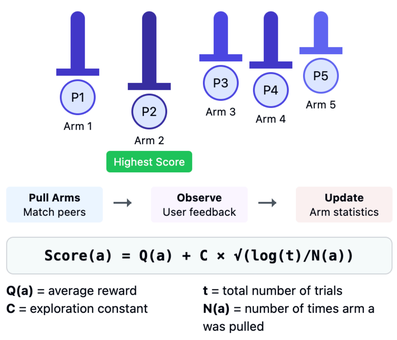

- Fine-tuned on 24,783 labeled examples
- Bidirectional context understanding
- Real-time moderation via FastAPI service
- Pattern-based enhancement for subtle toxicity

### Performance Metrics

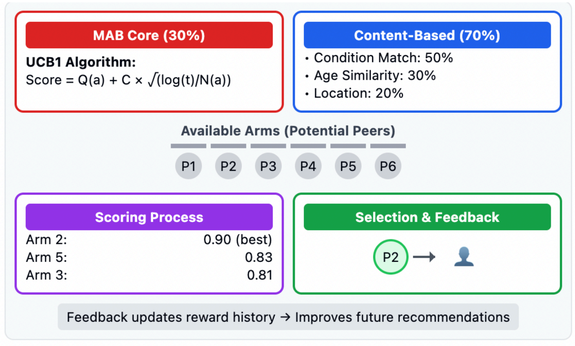
| Safe Content | Offensive | Hate Speech |
|---|---|---|
| 90% precision | 95% precision | 54% precision |
| 91% recall | 96% recall | 43% recall |
| F1: 90.5% | F1: 95.5% | F1: 48.0% |

## Peer Matching System

### Multi-Armed Bandit Architecture




### Hybrid Approach




### Cold Start Strategy

UCB1 strategy balances exploration and exploitation to optimize peer matches based on feedback

**Mass-follow simulation:**
Used faker.js to generate initial interaction patterns
Targeted 200 followers per user to establish initial data for the MAB algorithm

### Key Results

**Match Quality**
85.04% (Hybrid system)
vs. 26.4% comparison (baseline)

**Efficiency**
**0.12ms**
per match operation

## Future Work

- Improve hate speech detection capabilities
- Non-English content not currently supported
- Current approach doesn't account for treatment journey stage

### References

**Devlin, J., et al. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. NAACL.**
**Vaswani, A., et al. (2017). Attention is all you need. NeurIPS.**
**He, X., et al. (2017). Neural collaborative filtering. WWW Conference.**
**Yancey, K.P. & Settles, B. (2020). A sleeping, recovering bandit algorithm for optimizing recurring notifications. SIGKDD.**