

Augmenting Peer Mentor Emotional Intelligence for Digital Youth Support: Insights from Practice in Singapore

Arif Woozeer

Singapore Management University

In partnership with Over-The-Rainbow (OTR)

Advisor: Prof. Orlando Woods

Virtual Presentation



CASE STUDY & VIDEO
wooziecodes.github.io/otr-copilot
github.com/wooziecodes/otr-copilot

CONTEXT

Singapore youth increasingly seek mental health support via digital channels. OTR's volunteer peer mentors face significant challenges interpreting emotional cues in text-only, asynchronous conversations, without tools designed for this context.

APPROACH

This project developed an AI-augmented "copilot" system through iterative human-centred design, integrating psychological frameworks (RULER, Wheel of Emotions), culturally-adapted NLP, and Singapore English linguistics.

CONTRIBUTION

Design principles for AI integration in peer emotional support, extending Audrin & Audrin's digital emotional intelligence (dEI) framework with computational and collaborative dimensions. Augmentation, not automation.

BACKGROUND

The Digital Emotional Support Gap

Traditional emotional intelligence frameworks were built for face-to-face interaction. Digital peer support introduces fundamentally new interpretive challenges:

- OTR's Listeners — deliberately trained to be present, not prescriptive — provide support via real-time text chat without nonverbal cues
- Temporal ambiguity separates emotional content from context
- Singlish, code-switching, and emoji blends require culturally specific reading
- Cultural stigma shapes how distress is disclosed in Singapore youth
- No interpretive tools exist tailored to this peer support setting

Via COM-B, mentors face *capability* and *opportunity* gaps in reading digital cues; the copilot addresses both through enablement and environmental restructuring.

OTR deliberately avoids the term "peer supporter" — the Listener role is distinct; to hold space, not direct. This makes real-time AI intervention not merely disruptive but role-incompatible.

Digital Emotional Intelligence

Singlish / CSE

Peer Mentoring

COM-B

RESEARCH FRAMEWORK

Three Guiding Hypotheses

- H1 Augmentation: AI enhances interpretation accuracy**
A copilot combining RULER-grounded frameworks with computational NLP improves mentors' ability to read digital emotional cues compared to unaided interpretation.
- H2 Integration: workflow fit determines impact**
System effectiveness depends on seamless embedding within existing mentor workflows; even minor disruption negates benefit in emotionally sensitive conversations.
- H3 Complementary Intelligence: Human + AI > either alone**
The system enhances rather than replaces human judgment; AI provides pattern intelligence while humans retain the relational core of peer support.

METHODOLOGY

Human-Centred Design · 4 Phases

Iterative development with OTR stakeholders combining qualitative research, prototyping, and structured evaluation across 8 weeks.

01 DISCOVERY

Interviews, n=4 mentors; literature review across dEI, NLP, human-AI collaboration

02 DEFINITION

Affinity mapping, user journey maps, ethical impact assessment

03 DEVELOPMENT

3 iterative releases (Alpha → Beta v1 → Beta v2) with continuous stakeholder feedback

04 EVALUATION

Structured testing, n=5 OTR mentors; Likert scales + qualitative debrief

RULER Framework

Wheel of Emotions

CoSEM Corpus

VADER-adapted NLP

SYSTEM DESIGN: OTR COPILOT

Three Core Capabilities

- Digital Expression Pattern Recognition**
Monitors message timing, tone, and emoji usage adapted for Colloquial Singapore English: sentence-final particles, code-switching, reduplication patterns.
- Cross-Session Meta-Analysis**
Compares current messages against individualised baselines; surfaces recurring emotional patterns invisible in single-conversation views (AI's unique value proposition).
- AI-Assisted Response Templates**
Empathic and reflective suggestions via LLM, culturally calibrated for Singapore youth contexts, with explicit voice-preservation reminders to protect mentor authenticity.

FORMATIVE EVALUATION: PERFORMANCE METRICS

~65%

Emotion pattern accuracy vs. human annotation (formative prototype baseline)

3.4/5

Overall UX satisfaction from structured mentor testing

8

Distinct emotional categories mapped via Wheel of Emotions framework

HYPOTHESIS VALIDATION

- Partial** H1 **Augmentation:** AI adds value at session start; information overload limits sustained benefit. Real-time assistance overwhelms; cross-session meta-analysis enhances.
- Validated** H2 **Integration:** Workflow friction confirmed as the central adoption barrier. Minor disruption breaks emotional rapport in ways mentors could not recover from in-session.
- Validated** H3 **Complementary Intelligence:** Mentors retain full ownership of relationship and response; AI provides pattern intelligence that humans cannot replicate at scale.

KEY FINDING 1 · THE ASSISTANCE PARADOX

Temporal Framing Determines Whether AI Augments or Disrupts Emotional Presence

Mentors found real-time AI suggestions cognitively overwhelming during live support conversations. The same capabilities, presented as cross-session meta-analysis after the conversation, were highly valued and seen as irreplaceable. The temporal framing of AI assistance fundamentally changes its effect on mentor behaviour. For a Listener whose role is specifically to hold space rather than advise, concurrent AI output isn't merely a cognitive burden — it is antithetical to the function itself.

"The most effective AI support isn't concurrent; it's retrospective. AI should target temporally separable tasks where it offers comparative advantage, not tasks where human presence is the point."

DURING SESSION

Cognitive load ↑
Rapport breaks
Presence competes with output

AFTER SESSION

Emotional drift visible
Patterns surface
Mentor reviews freely

KEY FINDING 2 · LONGITUDINAL INTELLIGENCE

Cross-Session Pattern Detection Surfaces What Single-Session Observation Cannot

Mentors identified cross-session pattern detection as the system's irreplaceable capability: AI surfacing emotional drift and communication shifts across weeks that no individual supporter could track in isolation. This reframes the design question: not "how do we make real-time AI less intrusive?" but "how do we design AI for tasks temporally impossible for humans?"

"The strongest case for AI in peer support is not augmenting what mentors do in a session; it is doing what no mentor can do across sessions."

- 1 Single Session**
"What happened in this conversation?"
- 2 Cross-Session**
"What patterns are emerging over time?"
- 3 Longitudinal**
"What is invisible to any single mentor?"

KEY FINDING 3 · LLM IMPLEMENTATION RISK

RLHF-Trained Agreeableness Can Override System-Level Empathy Design

During prompt engineering, a consistent pattern emerged: when seekers communicated with emotionally avoidant language, Gemini's responses mirrored that avoidance rather than following the system-level empathy protocols in the prompt. The mechanism is RLHF-trained agreeableness: the model prioritises conversational tone-matching over explicit instruction. In clinical contexts this creates a quiet failure mode: the AI reinforces the avoidance patterns it was designed to gently challenge.

"LLM sycophancy is documented, but its clinical implication in emotional support contexts, where a model may silently mirror a seeker's avoidance rather than follow empathy protocols, remains underexplored." (cf. Sharma et al., 2023)

DESIGN PRINCIPLES · BEHAVIOUR CHANGE IMPLICATIONS

For AI in Emotional Support Contexts

Account for model disposition, not just model capability

Prompt engineering sets your clinical rules — but RLHF training sets the model's instincts. In emotional support contexts, those instincts can quietly override your rules. Audit for sycophantic drift before deployment, not after.

Cultural grounding is a clinical requirement

Standard NLP systematically misreads Singapore youth communication. Linguistic localisation is an ethical and clinical obligation, not a feature enhancement.

Dependency erosion risk is higher in volunteer contexts

Volunteer peer supporters lack the professional training to self-correct if AI scaffolding degrades their natural EI. AI in these settings must actively build human capability, not just support task performance.

THEORETICAL CONTRIBUTION

Extending dEI: and Where Next

- Adds *computational* and *collaborative* dimensions to Audrin & Audrin's (2023) digital emotional intelligence framework
- Operationalises the RULER framework computationally for asynchronous peer support
- Addresses the "assistance paradox" (Kumar et al., 2025) via coach-like AI that scaffolds rather than supplants mentor skill
- Identifies LLM sycophancy (Sharma et al., 2023) as a clinical implementation risk: model agreeableness can silently override system-level empathy protocols
- Culturally adapted NLP for Singlish/CSE using the CoSEM corpus (Gonzales et al., 2023)

OPEN QUESTIONS

- Does extended AI use erode mentor EI over time? Longitudinal study needed
- Suicide/self-harm risk alerting module with safety-first constraints
- Cross-session meta-analysis visualisations for supervisor use

