Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Adaptive stress-aware recommendation of supportive engagement activities for digital mental health: a contextual bandit approach with domain-specific language models

Ziinat Makaeva
Inje University
Muhammad Ahmed
Inje University
Md Ariful Islam Mozumder
Inje University
Sabeeha Jabeen
Inje University
Hee-Cheol Kim
heeki@inje.ac.kr

Inje University

Article

Keywords:

Posted Date: November 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11454113/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Adaptive stress-aware recommendation of supportive engagement activities for digital mental health: a contextual bandit approach with domain-specific language models

Makaeva Ziinat[1], Muhammad Ahmed[2], Md Ariful Islam Mozumder[2], Sabeeha Jabeen[2], and Hee-Cheol Kim[2,*]

[1]Department of Computer Engineering, Inje University, Republic of Korea
[2]Department of Digital Anti-Aging & Healthcare, Inje University, Republic of Korea
[*]heeki@inje.ac.kr

Abstract

Most digital mental-health tools give every user the same content, regardless of how that user feels now or what has helped them before. We built a pipeline that reads free text, detects stress and emotion with domain-adapted language models, and then picks a supportive activity with a contextual bandit. We also asked when per-user personalisation pays off. With five seeds per model, MentalRoBERTa outperformed retrained BERT-base and RoBERTa-base on stress detection (weighted F1 0.818). On our five-class, intervention-oriented emotion taxonomy it did not (macro F1 0.571). In a simulated engagement environment, simple non-contextual bandits did better than LinUCB at the population level. Per-user LinUCB gained nothing when it started from a shared population prior. Started from scratch, it beat the shared model by 0.179 in mean reward over 20 seeds, but only when simulated users had different activity preferences. We tested the classifiers on 70 free-text responses from real users, rated blind by a psychologist. Stress was detected well (area under the ROC curve 0.940). Emotion agreement was only fair (Cohen's κ 0.278), and most disagreements involved responses the psychologist labelled as overwhelm. The pipeline works technically, but its recommendation benefits come from simulation and need prospective testing.

Introduction

Stress and emotional distress place a heavy burden on public health; in 2019, anxiety disorders alone affected an estimated 301 million people worldwide[1]. Chatbots, mobile applications, and web-based programs can complement traditional care and reach many more people[2]. In a randomised controlled trial, the rule-based chatbot Woebot, which delivers cognitive behavioural therapy (CBT) content, reduced depressive symptoms[3], and studies of gamified digital interventions report mostly positive effects on mental-health outcomes[4]. Many of these tools, though, follow largely fixed content. They do not change what they offer when the user's mood changes, and they do not keep track of which activities helped that user before.

An adaptive tool has to do two things. It must recognise whether a user is distressed and in what way, and it must choose a response that is likely to help this particular user. For the first, Dreaddit[5] provides a benchmark of Reddit posts labelled for stress, on which fine-tuned BERT[6] and RoBERTa[7] reach approximately 76–82% F1[5;8]. Domain-adaptive pretraining, which continues

pretraining a language model on in-domain text[9], produced MentalBERT and MentalRoBERTa[8], which outperformed general-domain models on mental-health classification benchmarks. MentalRoBERTa has since served as the backbone of further task-specific models[10], and StressRoBERTa reports 81% F1 on Dreaddit[11]. Many of these comparisons report a single training run, so it is unclear how much of the reported advantage survives seed-to-seed variation. For emotion, GoEmotions[12] offers 27 emotion categories plus neutral, on which the original BERT baseline reaches 0.46 macro F1; recent work explores prompt design for emotion and mental-health recognition[13]. The GoEmotions authors also grouped the categories into Ekman's six basic emotions[12;14], but, to our knowledge, they have not been mapped onto a small taxonomy in which each class is defined by the supportive activity it should trigger.

For the second capability, contextual bandits learn online which action works best in a given context. In mobile health they are used in just-in-time adaptive interventions[15;16], which are commonly evaluated in micro-randomised trials[17;18]. Bandits have also been used to personalise physical-activity suggestions[19], and the DIAMANTE trial showed that they can personalise text-message interventions[20]. LinUCB[21;22] outperformed context-free bandits in news recommendation[21] and achieved lower regret than classical bandits in a recommender benchmark[23], but per-user personalisation typically needs enough interaction history before it beats a population model[24], and pooling information across users can speed up per-user learning in mobile health[25]. We are not aware of work that combines a domain-specific mental-health language model with bandit-based activity selection in one text-driven pipeline. We are also not aware of work that examines how initialising per-user models from a shared population prior affects their ability to learn individual preferences in such a pipeline.

Here we build and evaluate such a pipeline (Fig. 1) and address four questions. First, does a mental-health domain-adapted encoder outperform general-domain encoders on stress detection and on an intervention-oriented emotion taxonomy when all are trained under the same five-seed protocol? Second, how does LinUCB compare with simpler activity-selection policies in a simulated engagement environment? Third, when does per-user personalisation outperform a shared model, and how does starting from a shared population prior affect per-user learning? Fourth, how well do the classifiers work on free text written by real users and rated blind by a psychologist?

Results

Domain pretraining helps stress detection but not emotion classification

MentalRoBERTa achieved 0.818 ± 0.009 weighted F1 for stress detection on Dreaddit, significantly above BERT-base (0.790 ± 0.003, $p < 0.001$) and RoBERTa-base (0.803 ± 0.005, $p = 0.021$), all retrained under the same protocol (Table 1). For emotion classification, MentalRoBERTa reached 0.571 ± 0.011 macro F1 and 0.755 ± 0.013 weighted F1, statistically indistinguishable from both general-domain encoders on either metric ($p > 0.3$); RoBERTa-base's point estimate was marginally higher on weighted F1. The 95% confidence intervals of all three models overlapped substantially (Supplementary Table S1).

Because neutral makes up 77.6% of the emotion test set, we treat macro F1 as the primary emotion metric. Per-class F1 was 0.445 for anxiety, 0.473 for burnout, 0.481 for overwhelm, 0.620 for sadness, and 0.831 for neutral, with precision of only 0.31–0.38 for anxiety, burnout, and overwhelm (Supplementary Table S2). All three encoders showed the same error pattern (Fig. 2). Burnout and overwhelm were confused with each other in 12–17% of cases in both directions, and neutral was misclassified as burnout in 9–11%. Merging burnout and overwhelm into one class and retraining raised macro F1 from 0.571 ± 0.011 to 0.634 ± 0.010 (paired $t = 8.42$, $p < 0.001$). Because merging a

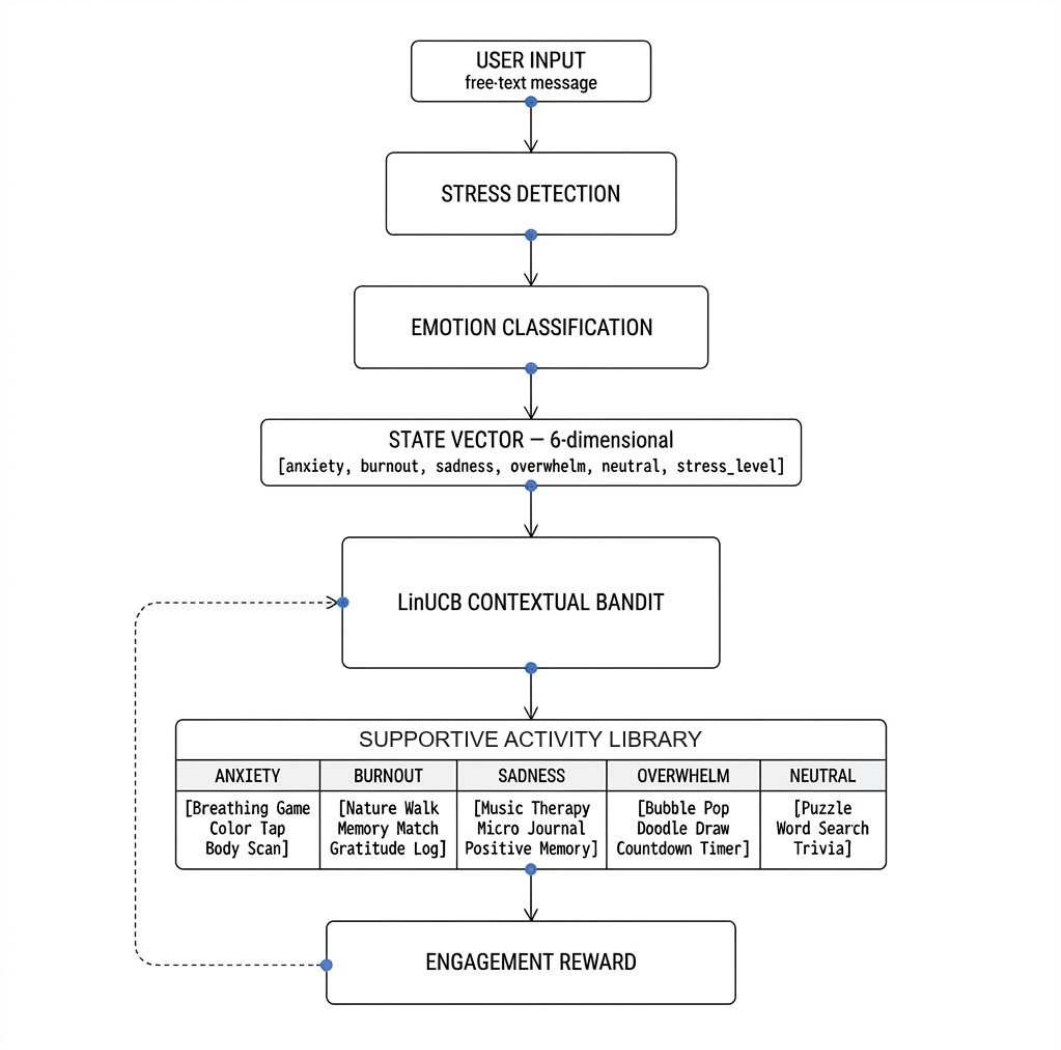


Figure 1: Pipeline overview. The stress model (MentalRoBERTa fine-tuned on Dreaddit) first screens each free-text message. Messages with a stress probability above 0.6 are passed to the emotion model (MentalRoBERTa fine-tuned on a five-class GoEmotions taxonomy). The detected emotion (one-hot) and the stress probability form a six-dimensional state, from which a LinUCB bandit chooses one of the three activities linked to that emotion. Simulated engagement provides the reward used to update the bandit.

confused pair mechanically removes their mutual errors, this gain does not indicate better emotion understanding; it shows that the burnout/overwhelm boundary is a major, encoder-independent source of error. All encoders were overconfident (expected calibration error, ECE, 0.091–0.103; Supplementary Fig. S1). In an independent rerun of the training protocol, temperature scaling reduced ECE by 35–43% without changing predictions (Supplementary Table S3).

Before settling on supervised fine-tuning, we tried two other approaches, and both did worse. Zero-shot LLaMA-3.2-3B, a large language model (LLM), got about 58% of emotions right on an informal test set, and its structured output was often inconsistent. K-means clusters built on linguistic features followed stress intensity rather than emotion type, with only 28% keyword-based emotion agreement (Supplementary Note 1). We also used an LLM judge (LLaMA-3.3-70B). It scored the supervised classifier's emotion output at 3.87 of 5 on 100 Dreaddit posts, compared with 2.81 for K-means, and scored the recommended activities at 4.14 of 5 on the 80 stressed posts

Table 1: Controlled encoder comparison (mean ± SD over 5 seeds). Stress: weighted F1 on the Dreaddit test set ($N = 715$). Emotion: five-class taxonomy on the single-label GoEmotions test set ($N = 4{,}590$). p-values are two-sided independent-samples t-tests against MentalRoBERTa, uncorrected for multiple comparisons.

	Stress		Emotion			
Model	Weighted F1	p	Macro F1	p	Weighted F1	p
BERT-base	0.790 ± 0.003	< 0.001	0.565 ± 0.012	0.453	0.746 ± 0.009	0.327
RoBERTa-base	0.803 ± 0.005	0.021	0.568 ± 0.010	0.695	0.756 ± 0.013	0.868
MentalRoBERTa	$\mathbf{0.818 \pm 0.009}$	–	0.571 ± 0.011	–	0.755 ± 0.013	–
StressRoBERTa[†]	0.810	n/a	–	–	–	–

[†]Published single-run value[11], not retrained under our protocol and not statistically comparable.

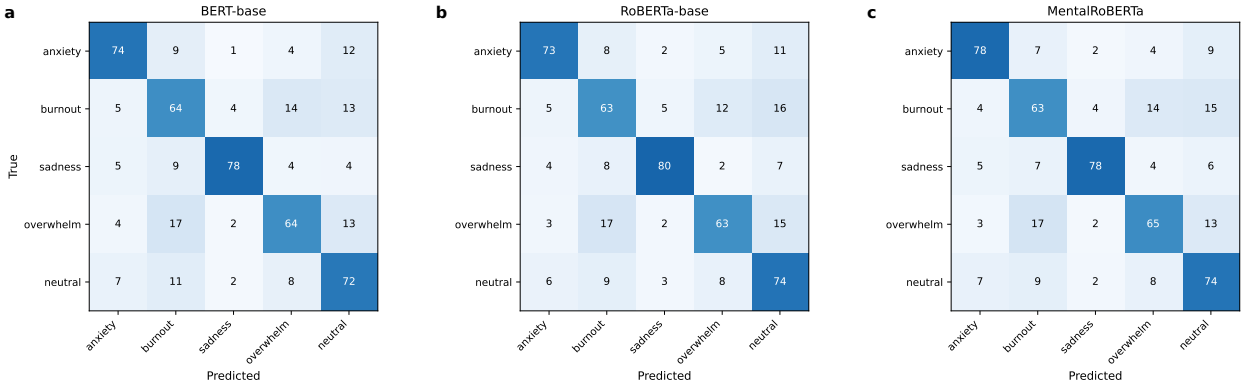


Figure 2: Emotion confusion matrices. Row-normalised confusion matrices (%), averaged over 5 seeds, for (**a**) BERT-base, (**b**) RoBERTa-base, and (**c**) MentalRoBERTa on the five-class GoEmotions test taxonomy. Rows are true classes and columns predicted classes. All three show the same dominant confusion between burnout and overwhelm.

(Supplementary Table S4).

In simulation, simple bandits outperform LinUCB at the population level

In the simulated engagement environment without per-user differences, LinUCB (mean reward 0.364 ± 0.049) significantly outperformed random selection (0.234 ± 0.012) and a rule-based policy that always selects the first activity listed for each emotion (0.320 ± 0.016), but was itself outperformed by ε-greedy (0.390 ± 0.018) and Thompson Sampling (0.398 ± 0.028) (Table 2). An oracle that knows the true engagement probabilities (0.427 ± 0.012) was never outperformed on any of the 20 seeds.

Table 2: Population-level activity selection in the simulated engagement environment (Methods; 20 seeds, 30 users, 100 sessions per user, no user heterogeneity). p-values are paired t-tests against LinUCB, Holm-corrected across four comparisons.

Policy	Mean reward	SD	Δ vs. LinUCB	Holm-corrected p
Random	0.234	0.012	-0.130	< 0.001
Rule-based	0.320	0.016	-0.044	0.002
LinUCB	0.364	0.049	–	–
ε-greedy	0.390	0.018	$+0.026$	0.012
Thompson Sampling	0.398	0.028	$+0.034$	0.007
Oracle	0.427	0.012	$+0.063$	n/a

These values follow analytically from the simulator. Ignoring noise, the expected reward of an activity with engagement probability p is $-0.7 + 1.482\,p^2$ (Eq. 3). The rule-based policy differs from the oracle only for sadness, where it chooses an activity with $p = 0.60$ instead of 0.85; this costs $1.482\,(0.85^2 - 0.60^2) = 0.537$ per sadness session, or 0.107 on average because sadness is one fifth of sessions, exactly the observed gap $(0.427 - 0.320)$. LinUCB's disadvantage most likely reflects the simulated task rather than the algorithm: reward does not depend on stress, and emotion is already used to restrict the candidate set, so LinUCB estimates a six-dimensional model per activity to recover what a per-activity mean already captures, and the extra parameters only add variance.

In simulation, per-user personalisation helps when not initialised from a shared prior

In the original configuration, both the shared and the per-user LinUCB bandits were initialised from a population model pretrained on 5,000 simulated episodes. Personalised LinUCB then showed no benefit over the shared model (paired $\Delta = -0.005$, $t = -1.002$, $p = 0.37$, 5 seeds; Table 3), and both remained far below the oracle that knows each user's preferences. The shared bandit's parameters moved about 4.5× farther from the prior than the average personalised bandit's (mean distance 0.104 vs. 0.023; per-seed ratio 4.0–5.8). With only 100 interactions per user, each personalised bandit stayed close to the population prior and so behaved like the shared model.

When the personalised bandits were started from scratch and the shared bandit kept its prior, personalised LinUCB outperformed the shared model on the five pre-specified seeds (paired $\Delta = +0.182$, $t = 9.000$, $p < 0.001$) and on 20 seeds ($\Delta = +0.179$, $t = 24.994$, $p = 5.4 \times 10^{-16}$, paired Cohen's $d = 5.59$, 95% CI $[0.164, 0.194]$; sign-flip permutation $p < 0.001$; Table 3). The effect persisted when both bandits were started from scratch ($\Delta = +0.219$, $p < 0.001$, 5 seeds), so it reflects personalisation rather than removal of the prior. The oracle remained above personalised LinUCB on every seed.

Table 3: Personalisation before and after correction in the simulated engagement environment (Methods; 30 users, 100 sessions per user, simulated user offset SD 0.30). Original: both bandits initialised from the shared prior, 5 seeds. Corrected: personalised bandits started from scratch, common random numbers, 20 seeds. Mean ± SD of mean reward.

Policy	Original	Corrected
Random	0.187 ± 0.013	0.197 ± 0.029
Population-best (fixed)	0.322 ± 0.038	0.327 ± 0.033
Shared LinUCB	0.329 ± 0.044	0.329 ± 0.034
Personalised LinUCB	0.325 ± 0.039	$\mathbf{0.508 \pm 0.021}$
Oracle (knows user offsets)	0.602 ± 0.007	0.599 ± 0.018
Personalised − shared (paired)	-0.005, $p = 0.37$	$+0.179$, $p = 5.4 \times 10^{-16}$

In the corrected configuration, personalised vs. shared, shared vs. random, and population-best vs. random remained significant after Holm correction; shared vs. population-best did not (raw $p = 0.41$).

Tracking reward over time showed that personalised LinUCB starts below the shared model while it explores and overtakes it after about 11 sessions (Fig. 3a). Its regret against the oracle fell from about 0.28 to 0.04 over 100 sessions, whereas the shared model's stayed at 0.25–0.29 (Fig. 3b). The gain depended on how much users differed (Supplementary Fig. S2). With identical simulated users, personalised LinUCB was significantly worse than the shared model ($\Delta = -0.118$, $p < 0.001$); the gain was positive at an offset SD of 0.15 ($\Delta = +0.045$, $p = 0.006$) and grew to $+0.348$ at 0.80. It remained positive across all tested exploration coefficients, noise levels, numbers of users, and sessions per user (all $p < 0.01$, 5 seeds; exploratory analysis).

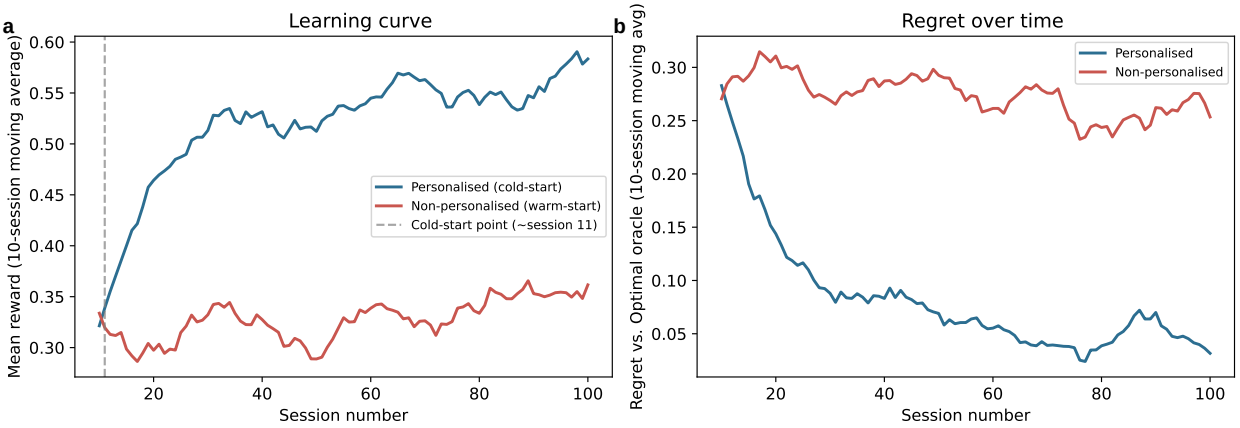


Figure 3: Learning curves for shared and personalised LinUCB. (**a**) Mean reward per session: personalised LinUCB (started from scratch) overtakes shared LinUCB (initialised from the population prior) after about 11 sessions. (**b**) Regret against the oracle decreases only for the personalised bandit. Simulated engagement environment, 10-session moving average, 5 seeds. In the legend, "non-personalised" denotes shared LinUCB.

Classification errors reduce simulated reward

When the shared bandit observed emotions sampled from MentalRoBERTa's empirical confusion matrix instead of the true emotion, mean reward fell from 0.360 to 0.096 ($p < 0.001$, 20 seeds), and

random emotion labels drove it to −0.366 (Supplementary Table S5). The magnitude depends on how strongly the simulator favours emotion-matched activities, but the direction is expected to hold for any reward that favours them. The routing threshold had a small effect. Across thresholds of 0.3–0.8, stress-class F1 stayed within 0.811–0.831, and the F1-maximising threshold of 0.5 exceeded the operating threshold of 0.6 by 0.007 F1 ($p = 0.047$), and the operating threshold routed 54.9% of messages to recommendation (Supplementary Table S6).

External validation on real-user text

Seventy adults (84% women, mostly students aged 18–34; Supplementary Table S7) described their current emotional state in free text and completed the Stress subscale of the 21-item Depression Anxiety Stress Scales (DASS-21). A psychologist, blind to all scores and model outputs, rated stress as present in 55 responses, absent in 10, and uncertain in 5. The stress classifier, applied without any fine-tuning, discriminated psychologist-rated stress well (area under the receiver operating characteristic curve, AUROC, 0.940, 95% CI [0.876, 0.987]; Table 4); at the 0.6 threshold, sensitivity was 0.855 (47/55) and specificity 1.000 (10/10), although specificity rests on only 10 stress-free responses. Discrimination of DASS-21 Stress dichotomised at its published normal cutoff was moderate (AUROC 0.717).

Table 4: External validation on real-user responses. AUROC CIs are bootstrap percentile intervals (5,000 resamples); κ is unweighted Cohen's κ with bootstrap CIs. Responses the psychologist rated uncertain were excluded from the comparisons with the psychologist. Self-report κ uses participants who chose a single taxonomy class. "First sentence only" applies the classifier to the first sentence of each response. "Merged" combines burnout and overwhelm.

Comparison	Statistic	95% CI	N
Stress classifier (AUROC)			
vs. psychologist (55 present / 10 absent)	0.940	[0.876, 0.987]	65
vs. DASS-21 Stress, normal cutoff (> 14)[26]	0.717	[0.573, 0.851]	70
vs. psychologist, first sentence only	0.973	–	65
Emotion classifier (Cohen's κ)			
vs. psychologist, five classes	0.278	[0.141, 0.415]	60
vs. psychologist, merged	0.323	[0.175, 0.468]	60
vs. psychologist, first sentence only	0.304	–	60
vs. self-report, five classes	0.234	[0.059, 0.426]	29
vs. self-report, merged	0.205	[0.025, 0.400]	29

Emotion agreement was only fair[27]: $\kappa = 0.278$ with the psychologist and 0.234 with self-report (Table 4). Merging burnout and overwhelm raised agreement with the psychologist to 0.323 but not with self-report, and all intervals overlapped. Agreement was higher for short responses (< 10 words: $\kappa = 0.525$, $N = 16$) than for long ones ($\kappa = 0.193$, $N = 44$). Disagreement concentrated in longer, multi-clause narratives that the psychologist coded as overwhelm and the classifier resolved into a single emotion, most often sadness or anxiety. Truncating responses to their first sentence changed agreement only slightly, so the mismatch between short training comments and longer survey narratives is a minor factor. The definition of the overwhelm class, which was trained on the GoEmotions categories anger and annoyance (Table 5), appears to contribute most to the disagreement.

Discussion

Domain-adaptive pretraining improved binary stress detection but not five-class emotion classification, with both tested under the same controlled protocol. In simulated activity selection, added model capacity helped only when the environment supplied information the model could use: LinUCB lost to simpler bandits when emotion already restricted the candidate set, and per-user models helped only when users differed and enough interactions per user accumulated, consistent with the cold-start literature[24]. Stress detection also agreed well with a psychologist's reading of real users' descriptions of their state, although only moderately with self-reported DASS-21 stress, whereas the emotion taxonomy did not transfer.

One explanation for the pretraining result is that pretraining on mental-health communities[8] sharpens sensitivity to *whether* clinically relevant distress is present, which is the core of stress detection, more than to *which* of several adjacent distress states a passage expresses. General-domain pretraining may capture this second distinction equally well. Closer vocabulary overlap between the mental-health pretraining corpus and Dreaddit, both drawn from Reddit communities, may also contribute. Because both tasks were evaluated with the same five seeds, data handling, and hyperparameters, the contrast is unlikely to be an artefact of seed variation, which single-run comparisons cannot rule out.

The initial null personalisation result reflected a design choice rather than an inability to learn individual differences: per-user models were initialised from a population prior whose weight (5,000 episodes) dominated a 100-interaction budget. Once that constraint was removed, LinUCB learned persistent per-user preferences. The sensitivity analysis suggests a hypothesis for deployment that can be tested with real users: a per-user bandit started from scratch should help when real users differ in their activity preferences, and should harm engagement when they do not. Real inter-user variation should therefore be measured before choosing between the two architectures. The two configurations also differ in the use of common random numbers; the check with both bandits started from scratch holds initialisation fixed but does not remove this difference.

The external validation supports the stress component for further prospective testing. Agreement with DASS-21 (AUROC 0.717) was lower than with the psychologist (0.940), as expected, because the classifier and the psychologist judged the same text, whereas DASS-21 measures self-reported symptoms over the past week. For the emotion component, the validation suggests a specific revision. The overwhelm class, trained on GoEmotions anger and annoyance, could be redefined to match how it is applied to multi-clause narratives, supported by clearer guidelines separating burnout from overwhelm and more labelled examples at that boundary. Because classification errors reduced simulated reward by about three quarters, simulated reward appears to be limited more by classification accuracy than by bandit tuning.

Methods

Datasets

Dreaddit[5] contains 3,553 crowd-annotated Reddit segments from five stress-prone domains, labelled stress or no stress; we used the official split of 2,838 training (52.4% stress) and 715 test samples. GoEmotions[12] contains about 58,000 English Reddit comments labelled with 27 emotion categories plus neutral; its official split (43,410 training, 5,426 validation, 5,427 test) contains the 54,263 comments retained after the dataset authors' filtering for rater agreement. We kept only comments with a single gold label (excluding 7,102 training, 878 validation, and 837 test comments), leaving 36,308 training, 4,548 validation, and 4,590 test comments. The test set (174 anxiety, 382 burnout,

148 sadness, 325 overwhelm, 3,561 neutral) was used throughout.

Emotion taxonomy and activities

We defined five classes, each linked to an established supportive technique, and mapped the GoEmotions categories onto them (Table 5). Each class has three brief, low-intensity, game-based activities, 15 in total, intended as short engagement tools rather than stand-alone treatments. For example, the breathing exercise uses a 4-7-8 pattern[28], the nature walk offers passive visual restoration[29], and the positive-memory activity draws on benefit-finding writing[30]. These sources informed the groupings; the therapeutic effect of the activities themselves was not tested.

Table 5: Emotion taxonomy and simulator parameters. For each class: theoretical basis, mapped GoEmotions categories, the three candidate activities with their assumed base engagement probabilities $\mu_{e,a}$ (the first listed is chosen by the rule-based policy), and the range from which a session's stress level is drawn. $\mu_{e,a}$ are assumptions of the simulator, not estimates from user data.

Class	Basis	GoEmotions categories	Activities ($\mu_{e,a}$)	Stress range
Anxiety	CBT, diaphragmatic breathing[31]	fear, nervousness, confusion	Breathing game (0.90), colour tap (0.75), body scan (0.70)	$[0.7, 1.0]$
Burnout	Attention restoration[29]	disgust, disapproval, embarrassment, disappointment	Nature walk (0.90), memory match (0.75), gratitude log (0.80)	$[0.6, 0.9]$
Sadness	Expressive writing[32]	sadness, grief, remorse	Music therapy (0.60), micro journal (0.75), positive memory (0.85)	$[0.5, 0.9]$
Overwhelm	Flow theory[33]	anger, annoyance	Bubble pop (0.85), doodle draw (0.80), countdown timer (0.75)	$[0.7, 1.0]$
Neutral	Engagement maintenance	all remaining, including positive emotions such as joy, love, and admiration	Puzzle (0.85), word search (0.80), trivia (0.75)	$[0.0, 0.4]$

Classification models and training

Our main encoder was `mental/mental-roberta-base`[8], a RoBERTa-base model that had been pretrained further on posts from mental-health communities on Reddit. We fine-tuned it twice, once for binary stress detection and once for the five emotion classes. BERT-base[6] and RoBERTa-base[7] went through exactly the same training. The emotion data were unbalanced, so we undersampled them by class. Sadness was the smallest class (1,209 posts); neutral was capped at three times that number and every other class at two times, which left 11,045 training samples. This set was drawn once with a fixed seed, and all encoders and seeds used the same copy. Each model was trained with five seeds (42, 123, 456, 789 and 2024). The optimiser was AdamW with a learning rate of 2×10^{-5} and weight decay of 0.01. We used batches of 16, sequences of up to 128 tokens and gradient clipping at 1.0. Emotion models ran for 5 epochs. All runs were done on a single NVIDIA GeForce GTX 1070 (8 GB) with Python 3.12.1, PyTorch 2.6.0 (CUDA 12.4), Hugging Face Transformers 5.8.0, scikit-learn 1.8.0, NumPy and SciPy.

To measure calibration, we pooled the seeds and computed the ECE. Temperature scaling[34] was fitted for each seed on the GoEmotions validation split. The numbers for temperature scaling were taken from a separate rerun of the same training protocol (Supplementary Table S3). In the taxonomy ablation, burnout and overwhelm became one class, and MentalRoBERTa was trained

again with the same settings. We also looked at two other ways of encoding emotion. One was LLaMA-3.2-3B used zero-shot; this model had been QLoRA-tuned[35] on MentalChat16K but had never seen our labels. The other was K-means with $K = 5$ on nine Linguistic Inquiry and Word Count (LIWC) features and the stress probability (Supplementary Note 1). Finally, LLaMA-3.3-70B served as a judge[36–38]. It scored emotion-detection accuracy and how appropriate the activity was, each on a 1–5 scale, for 100 Dreaddit posts (20 for each emotion class). These scores were tested against 3.0 with one-sample t-tests.

Contextual bandit

A message is routed to recommendation if its stress probability exceeds 0.6. LinUCB[21] then uses the state $\mathbf{s} = [e_{\text{anx}}, e_{\text{burn}}, e_{\text{sad}}, e_{\text{over}}, e_{\text{neu}}, \sigma]^\top \in \mathbb{R}^6$, where e_i one-hot encodes the detected emotion and σ is the stress probability, and selects among the three activities $\mathcal{A}(e)$ assigned to that emotion:

$$a^* = \arg\max_{a \in \mathcal{A}(e)} \boldsymbol{\theta}_a^\top \mathbf{s} + \alpha\sqrt{\mathbf{s}^\top \mathbf{A}_a^{-1}\mathbf{s}}, \qquad \boldsymbol{\theta}_a = \mathbf{A}_a^{-1}\mathbf{b}_a, \tag{1}$$

with $\mathbf{A}_a = \mathbf{I}$, $\mathbf{b}_a = \mathbf{0}$ initially and $\alpha = 0.5$. After reward r, $\mathbf{A}_a \leftarrow \mathbf{A}_a + \mathbf{s}\mathbf{s}^\top$ and $\mathbf{b}_a \leftarrow \mathbf{b}_a + r\,\mathbf{s}$. The population bandit was pretrained on 5,000 synthetic episodes, generated afresh for each seed, pairing an emotion and stress level with a random activity and a simulated reward.

Simulated engagement and reward

For user u, emotion e, and activity a, the engagement probability is $p = \text{clip}(\mu_{e,a} + \delta_{u,a} + \varepsilon, 0, 1)$, with base value $\mu_{e,a}$ (Table 5), persistent per-user offset $\delta_{u,a}$, and noise $\varepsilon \sim \mathcal{N}(0, 0.08^2)$. A click occurs with probability p; without a click, playtime $T = 0$. After a click, a second independent draw with probability p decides between sustained play ($T \sim \text{Uniform}(0.5, 15)$ minutes) and an early exit ($T \sim \text{Uniform}(0, 0.5)$). A session is a bounce if there is no click or $T < 0.5$, and

$$R = \begin{cases} -0.7 & \text{if bounce,} \\ 0.3 + 0.7\min(1,\, T/10) & \text{otherwise.} \end{cases} \tag{2}$$

Unit tests confirmed that $R \in \{-0.7\} \cup [0.3, 1.0]$ and that an unopened activity never receives positive reward (Supplementary Note 2). Ignoring noise, sustained play occurs with probability p^2 and has mean reward 0.782, so

$$\mathbb{E}[R \mid p] = -0.7 + (0.7 + 0.782)\,p^2 \approx -0.7 + 1.482\,p^2. \tag{3}$$

The simulator assumes that expected reward depends on the true emotion, the activity, and a persistent per-user offset but not on stress level, and that each session's emotion is drawn uniformly from the five classes and its stress level from the class-specific range. The stress feature therefore carries no reward signal in the simulation; the effect of user heterogeneity is examined by varying the offset SD in the sensitivity analysis.

Policies and personalisation experiments

Population-level experiments compared random selection, the rule-based policy (first-listed activity, no learning; Supplementary Note 3), LinUCB, ε-greedy[39] ($\varepsilon = 0.1$), Gaussian Thompson Sampling[40;41], and an oracle, with $\delta_{u,a} = 0$. For personalisation, each simulated user received offsets $\delta_{u,a} \sim \mathcal{N}(0, 0.30^2)$, fixed across that user's sessions, and each seed drew a fresh set of 30

users with 100 sessions each. We compared random selection; population-best, the activity with the highest $\mu_{e,a}$; a shared LinUCB bandit; one LinUCB bandit per user; and an oracle that knows each user's offsets,

$$\pi^*(e,u) = \arg\max_{a \in \mathcal{A}(e)} \left(\mu_{e,a} + \delta_{u,a}\right), \tag{4}$$

which reduces to population-best when users do not differ. In the original configuration, both bandits were cloned from the pretrained population bandit; in the corrected configuration, only the per-user bandits were started from scratch. Parameter displacement was measured as $\|\boldsymbol{\theta}_a^{\text{final}} - \boldsymbol{\theta}_a^{\text{prior}}\|_2$, averaged over activities and, for the personalised model, over users. In the corrected configuration, one noise realisation was pre-drawn per user, session, and activity and reused by every policy choosing that activity (common random numbers[42]). All configurations are listed in Supplementary Table S8, and the reasons for reporting both personalisation configurations are given in Supplementary Note 4. For the error-propagation analysis, the shared bandit observed the true emotion, an emotion sampled from MentalRoBERTa's empirical confusion matrix, or a random emotion, while reward always followed the true emotion. The exploratory sensitivity analysis varied one setting at a time around the corrected configuration: $\alpha \in \{0.1, 0.3, 0.5, 0.8, 1.2\}$, noise SD $\in \{0.04, 0.08, 0.15, 0.25\}$, users $\in \{10, 30, 50, 100\}$, offset SD $\in \{0, 0.15, 0.30, 0.50, 0.80\}$, and sessions per user $\in \{50, 100, 200, 400\}$.

External validation study

To see how the classifiers behave outside Reddit, we ran a short anonymous survey in English between 22 September and 2 October 2026. Nobody was asked for a name or an e-mail address. Free-text answers that might still reveal who wrote them were looked for by hand. In the first question, people described in their own words how they felt right then. After that came the seven-item DASS-21 Stress subscale[26], a stress rating from 0 to 10 on a visual analogue scale, and a question asking which of the five emotions fitted them best (*other/unsure* was also possible). S.J., a psychologist, read the free text and nothing else. She marked stress as present, absent or *uncertain* and named the primary emotion, without seeing any questionnaire score, self-report or model output. No fine-tuning was done on these data and the 0.6 threshold stayed as it was. Both models had been saved before the survey started: the prototype stress model on 13 May 2026 (seed 42, after epoch 2) and the five-class emotion model on 20 August 2026.

Statistical analysis

For the classifiers we report the mean ± SD of the five seeds and use two-sided independent-samples t-tests to compare models. The bandit experiments were analysed differently. Every seed draws a new set of users, and with common random numbers all policies inside one seed see the same draws, so we compared seed-level means with paired t-tests. A sign-flip permutation test with 10,000 permutations was run as a check on these results. Each experiment had four pairwise comparisons, and we applied the Holm–Bonferroni correction to them. Per-user observations were never pooled. Users who interact with the same shared bandit are not independent of each other. In the sensitivity analysis we tested per-seed differences with one-sample t-tests and did not correct for multiplicity. In the external validation, AUROC confidence intervals come from 5,000 bootstrap resamples. For the comparison with the psychologist we also give DeLong's interval ($[0.853, 0.977]$). Confidence intervals for unweighted κ were bootstrapped as well. We split responses by length (< 10 words) only after seeing the data. DASS-21 Stress was dichotomised at the published normal cutoff (doubled score > 14), which gave 43 elevated and 27 normal scores. Responses that the psychologist rated uncertain were left out.

LLM support in the research workflow

During preparation of this work, the authors used Claude (Anthropic) for language editing, manuscript organisation, and code review, and Grammarly for language checks. No generative AI was used to generate data, model outputs, statistical results, or conclusions. The authors reviewed and edited all AI-assisted material, verified the references and reported values against the original sources and stored results, and take full responsibility for the content of the article.

Limitations

The main limitation is that every recommendation result comes from simulation. The reward table and the level of user heterogeneity were assumed, not estimated from user data, and the full loop was not tested prospectively, so we make no claim of clinical benefit. Two labels are also imperfect: our burnout class covers more general negative affect than occupational burnout, and Dreaddit's "no stress" label does not mean the same as our neutral class. Weighted emotion F1 mostly reflects the neutral class, and the calibration ranking of the encoders changed between independent runs. The models saw each message on its own, while a deployed system would build up context over many sessions. The external validation was small (70 participants, mostly young women and students) and had one psychologist, so we could not estimate inter-rater reliability. Finally, we did not retrain StressRoBERTa under our protocol.

The system is meant to support users, not to replace professional care. Our prototype has no crisis detection or escalation, so a high-stress message goes to activity recommendation like any other. Before any use with real users, a crisis screen must come first and direct such users to crisis resources instead of the recommender. Next steps are a larger user study with at least two raters, which would also let us measure real per-user heterogeneity; a revised emotion taxonomy; a comparison of per-user and shared bandits, including per-user ε-greedy and Thompson Sampling, on real engagement data; and memory of earlier sessions.

Conclusions

Domain-specific language models and a contextual bandit can be joined in one pipeline that recommends short supportive activities from free text. For stress, MentalRoBERTa was the strongest of the three encoders in our five-seed comparison (weighted F1 0.818). For the five emotion classes it was no better than BERT-base or RoBERTa-base (macro F1 0.571), and much of the remaining error lies between burnout and overwhelm. In the simulation, simple non-contextual bandits outperformed LinUCB at the population level. Per-user LinUCB did better than the shared model only when it started from scratch and the simulated users actually differed in the activities they preferred. On text from real users, the stress classifier matched the psychologist's judgement well (AUROC 0.940). Emotion agreement stayed at a fair level (Cohen's κ 0.278). We therefore consider stress detection ready for prospective testing; the emotion taxonomy should be revised first. So far the benefit of the recommendations has been shown only in simulation, and a prospective study with real users is needed before any therapeutic claim can be made.

References

[1] Yang, X. *et al.* Global, regional and national burden of anxiety disorders from 1990 to 2019: results from the Global Burden of Disease Study 2019. *Epidemiol. Psychiatr. Sci.* **30**, e36 (2021).

[2] Abd-Alrazaq, A. A., Rababeh, A., Alajlani, M., Bewick, B. M. & Househ, M. Effectiveness and safety of using chatbots to improve mental health: systematic review and meta-analysis. *J. Med. Internet Res.* **22**, e16021 (2020).

[3] Fitzpatrick, K. K., Darcy, A. & Vierhile, M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. *JMIR Ment. Health* **4**, e19 (2017).

[4] Aschentrup, L. *et al.* Effectiveness of gamified digital interventions in mental health prevention and health promotion among adults: a scoping review. *BMC Public Health* **24**, 69 (2024).

[5] Turcan, E. & McKeown, K. Dreaddit: a Reddit dataset for stress analysis in social media. In *Proc. 10th International Workshop on Health Text Mining and Information Analysis (LOUHI)* 97–107 (2019).

[6] Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. BERT: pre-training of deep bidirectional transformers for language understanding. In *Proc. NAACL-HLT 2019* 4171–4186 (2019).

[7] Liu, Y. *et al.* RoBERTa: a robustly optimized BERT pretraining approach. Preprint at https://arxiv.org/abs/1907.11692 (2019).

[8] Ji, S. *et al.* MentalBERT: publicly available pretrained language models for mental healthcare. In *Proc. 13th Language Resources and Evaluation Conference* 7184–7190 (2022).

[9] Gururangan, S. *et al.* Don't stop pretraining: adapt language models to domains and tasks. In *Proc. 58th Annual Meeting of the Association for Computational Linguistics* 8342–8360 (2020).

[10] Wagay, F. A., Jahiruddin & Altaf, Y. MentalRoBERTa-Caps: a capsule-enhanced transformer model for mental health classification. *MethodsX* **15**, 103483 (2025).

[11] Alqahtani, A. A., Kayi, E. & Diab, M. T. StressRoBERTa: cross-condition transfer learning from depression, anxiety, and PTSD to stress detection. In *Proc. 1st Workshop on Linguistic Analysis for Health (HeaLing 2026)* 305–313 (Association for Computational Linguistics, 2026).

[12] Demszky, D. *et al.* GoEmotions: a dataset of fine-grained emotions. In *Proc. 58th Annual Meeting of the Association for Computational Linguistics* 4040–4054 (2020).

[13] Kim, M., Kim, T., Vo, T. H. A., Jung, Y. & Lee, U. Exploring modular prompt design for emotion and mental health recognition. In *Proc. 2025 CHI Conference on Human Factors in Computing Systems* (2025).

[14] Ekman, P. An argument for basic emotions. *Cogn. Emot.* **6**, 169–200 (1992).

[15] Nahum-Shani, I. *et al.* Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. *Ann. Behav. Med.* **52**, 446–462 (2018).

[16] Tewari, A. & Murphy, S. A. From ads to interventions: contextual bandits in mobile health. In *Mobile Health: Sensors, Analytic Methods, and Applications* 495–517 (Springer, 2017).

[17] Liao, P., Klasnja, P., Tewari, A. & Murphy, S. A. Sample size calculations for micro-randomized trials in mHealth. *Stat. Med.* **35**, 1944–1971 (2016).

[18] Klasnja, P. *et al.* Efficacy of contextually tailored suggestions for physical activity: a micro-randomized optimization trial of HeartSteps. *Ann. Behav. Med.* **53**, 573–582 (2019).

[19] Liao, P., Greenewald, K., Klasnja, P. & Murphy, S. Personalized HeartSteps: a reinforcement learning algorithm for optimizing physical activity. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **4**, 18 (2020).

[20] Aguilera, A. *et al.* Effectiveness of a digital health intervention leveraging reinforcement learning: results from the Diabetes and Mental Health Adaptive Notification Tracking and Evaluation (DIAMANTE) randomized clinical trial. *J. Med. Internet Res.* **26**, e60834 (2024).

[21] Li, L., Chu, W., Langford, J. & Schapire, R. E. A contextual-bandit approach to personalized news article recommendation. In *Proc. 19th International Conference on World Wide Web* 661–670 (2010).

[22] Chu, W., Li, L., Reyzin, L. & Schapire, R. E. Contextual bandits with linear payoff functions. In *Proc. 14th International Conference on Artificial Intelligence and Statistics* 208–214 (2011).

[23] Li, Z. Comparative analysis of classical and LinUCB bandits in recommender systems based on cumulative regret and reward. *ITM Web Conf.* **78**, 01031 (2025).

[24] Silva, N., Silva, T., Werneck, H., Rocha, L. & Pereira, A. User cold-start problem in multi-armed bandits: when the first recommendations guide the user's experience. *ACM Trans. Recomm. Syst.* **1**, 1–24 (2023).

[25] Tomkins, S., Liao, P., Klasnja, P. & Murphy, S. IntelligentPooling: practical Thompson sampling for mHealth. *Mach. Learn.* **110**, 2685–2727 (2021).

[26] Lovibond, P. F. & Lovibond, S. H. The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behav. Res. Ther.* **33**, 335–343 (1995).
[27] Landis, J. R. & Koch, G. G. The measurement of observer agreement for categorical data. *Biometrics* **33**, 159–174 (1977).
[28] Zaccaro, A. *et al.* How breath-control can change your life: a systematic review on psycho-physiological correlates of slow breathing. *Front. Hum. Neurosci.* **12**, 353 (2018).
[29] Kaplan, S. The restorative benefits of nature: toward an integrative framework. *J. Environ. Psychol.* **15**, 169–182 (1995).
[30] Pennebaker, J. W. & Beall, S. K. Confronting a traumatic event: toward an understanding of inhibition and disease. *J. Abnorm. Psychol.* **95**, 274–281 (1986).
[31] Barlow, D. H. *Anxiety and Its Disorders: The Nature and Treatment of Anxiety and Panic* 2nd edn (Guilford Press, 2002).
[32] Pennebaker, J. W. Writing about emotional experiences as a therapeutic process. *Psychol. Sci.* **8**, 162–166 (1997).
[33] Csikszentmihalyi, M. *Flow: The Psychology of Optimal Experience* (Harper & Row, 1990).
[34] Guo, C., Pleiss, G., Sun, Y. & Weinberger, K. Q. On calibration of modern neural networks. In *Proc. 34th International Conference on Machine Learning* 1321–1330 (2017).
[35] Dettmers, T., Pagnoni, A., Holtzman, A. & Zettlemoyer, L. QLoRA: efficient finetuning of quantized LLMs. *Adv. Neural Inf. Process. Syst.* **36** (2023).
[36] Zheng, L. *et al.* Judging LLM-as-a-judge with MT-Bench and Chatbot Arena. *Adv. Neural Inf. Process. Syst.* **36** (2023).
[37] Liu, Y. *et al.* G-Eval: NLG evaluation using GPT-4 with better human alignment. In *Proc. 2023 Conference on Empirical Methods in Natural Language Processing* 2511–2522 (2023).
[38] Croxford, E. *et al.* Automating evaluation of AI text generation in healthcare with a large language model (LLM)-as-a-judge. Preprint at https://doi.org/10.1101/2025.04.22.25326219 (2025).
[39] Sutton, R. S. & Barto, A. G. *Reinforcement Learning: An Introduction* 2nd edn (MIT Press, 2018).
[40] Chapelle, O. & Li, L. An empirical evaluation of Thompson sampling. *Adv. Neural Inf. Process. Syst.* **24**, 2249–2257 (2011).
[41] Russo, D. J., Van Roy, B., Kazerouni, A., Osband, I. & Wen, Z. A tutorial on Thompson sampling. *Found. Trends Mach. Learn.* **11**, 1–96 (2018).
[42] Law, A. M. & Kelton, W. D. *Simulation Modeling and Analysis* 3rd edn (McGraw-Hill, 2000).

Ethics statement

The main experiments used the public, anonymised Dreaddit and GoEmotions datasets, which contain no personal information. The external validation survey was fully anonymous and collected no personally identifying information. Participation was voluntary, and all 70 participants gave informed consent before answering any question.

Data availability

Dreaddit and GoEmotions are publicly available[5;12]. The external validation survey data contain free-text self-reports of emotional state and are not publicly released, in line with the anonymity and consent terms under which they were collected. The GoEmotions-to-class mapping, simulator, bandit and personalisation code, and training code are available at https://github.com/ZiiMceva/mental-health-chatbot and archived at Zenodo (https://doi.org/10.5281/zenodo.23488101).

Acknowledgements

This work was supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) - Innovative Human Resource Development for Local Intellectualization program grant funded by the Korea government (MSIT) (IITP-2024-RS-2024-00436773).

Funding

This research was funded by the Institute of Information & Communications Technology Planning & Evaluation (IITP) - Innovative Human Resource Development for Local Intellectualization program grant funded by the Korea government (MSIT) (IITP-2024-RS-2024-00436773).

Author contributions

M.Z. conceived the study, developed the software, performed the experiments, analysed the data and wrote the manuscript. M.A. and M.A.I.M. contributed to the methodology and validation and reviewed and edited the manuscript. S.J., a psychologist, performed the blind expert rating of the external validation responses and reviewed and edited the manuscript. H.-C.K. supervised the project and reviewed and edited the manuscript. All authors reviewed and approved the final manuscript.

Additional information

Supplementary information accompanies this paper (Supplementary Tables S1–S8, Supplementary Figs. S1–S2, and Supplementary Notes 1–4).
Competing interests: The authors declare no competing interests.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation.pdf