EAI Endorsed Transactions
on Pervasive Health and Technology

Research Article EAI.EU

Easy Assistant: An App Design Study to Alleviate Daily Work Anxiety Among College Faculty

Jiao Jiang[1], Haoteng Chen[1*], Yousheng Yao[1,2,3*], Hao Gu[1], Junfeng Chen[1] and Yesheng Hong[3,4]

[1]Hexiangning College of Art and Design, Zhongkai University of Agriculture and Engineering, No. 24, Dongsha Street, Haizhu District, Guangzhou City, Guangdong Province, China
[2]Faculty of Humanities and Arts, Macau University of Science and Technology, Avenida Wai Long, Taipa, Macau, China
[3]People's Government of Jiangdong Town, No. 1 Fuqian Street, Jiangdong Town, Chaoan District, Chaozhou City, Guangdong Province, China
[4]Jiangdong Town Industrial Development Service Center, Chaoan District, Chaozhou City, Guangdong Province, China

Abstract

INTRODUCTION: Faculty juggle teaching, research, administration, and student support amid unstructured email, chat, and LMS alerts. These demands are linked to increased extraneous cognitive load and are associated with work anxiety. We evaluate a CLT-guided NLP app that structures messages into tasks.
OBJECTIVES: Assess technical performance and preliminary pre–post changes in anxiety and workload.
METHODS: Stage I validated a four-stage pipeline (cleaning; intent/entity recognition; probability calibration (evaluated by ECE); task structuring with priority gating). Stage II was a 4-week single-group pre–post feasibility study ($N = 30$) using GAD-7 and NASA-TLX.
RESULTS: Intent F1 = 0.92; entity F1 = 0.89; 87% of tasks met TC $\geq$ 0.8. GAD-7 decreased by 2.7 points (95% CI [−3.6, −1.8]) and NASA-TLX by 16.3 (95% CI [−20.5, −12.1]); usability was high (SUS = 78.2).
CONCLUSION: We observed pre–post reductions in workload and anxiety after four weeks of use; causal inference awaits controlled trials.

Keywords: Cognitive Load Theory; Natural Language Processing; Occupational mental health; Work anxiety; Cognitive load; mHealth

Received on 22 November 2025, accepted on 06 January 2026, published on 15 January 2026

doi: 10.4108/ eetpht.11.11052

1. Introduction

Faculty in higher education often work in fragmented and cognitively demanding environments, juggling teaching, research, administration, and mentoring. These conditions are closely linked to elevated stress and a need for better tools to manage workload and well-being [1,2]. The difficulty lies not only in "too much to do," but also in "too much, arriving in the wrong form."

Mobile mental-health apps have emerged as a key tool for psychological support, with studies showing reductions in stress and improvements in psychosocial outcomes [3,4].

These tools often use brief, structured activities delivered on personal devices to help regulate distress at scale [5,6]. However, most existing apps focus on post-hoc coping strategies, such as mindfulness and relaxation techniques, rather than addressing the cognitive structure of work-related stress. These strategies fail to reduce the underlying cognitive burden suggest by fragmented communication. Our approach, grounded in Cognitive Load Theory (CLT), focuses on proactively managing cognitive load by transforming unstructured messages into actionable tasks,

* Corresponding author. Email: 2978277328@qq.com ; yaoyousheng41@gmail.com

reducing extraneous load and aligning work with faculty's academic goals[7]. For instructors whose day is organized by their inbox, reducing this information-to-task burden is as critical as providing coping exercises.

We combine CLT with the Job Demands–Resources (JD-R) model to offer a more comprehensive understanding of faculty work anxiety. This integration is a key innovation of this study, offering a framework that not only reduces cognitive overload through CLT but also contextualizes the anxiety within the broader scope of job demands and available resources. CLT explains how unstructured communication increases cognitive load, while JD-R contextualizes this within broader job demands and resources. Our intervention, Easy Assistant, uses NLP to convert unstructured faculty communications into prioritized task cards, reducing cognitive load directly in the communication stream, without requiring faculty to engage with another wellness app after hours.

This intervention shifts the focus from post-hoc coping to proactive task management. Conceptually, our contribution is threefold (Table 1). First, at the theoretical level, we integrate Cognitive Load Theory (CLT) with the Job Demands–Resources (JD–R) framework and emotion-labor research to frame unstructured digital messages as a specific job demand for faculty. Second, at the methodological level, we operationalize this framework through an end-to-end NLP pipeline and a CLT-guided interface that together restructure messages into prioritized tasks with explicit uncertainty handling. Third, at the application level, we target college faculty members' everyday digital communication rather than generic productivity workflows or broad mental-health support, aligning the intervention tightly with their occupational context. In doing so, Easy Assistant directly targets the cognitive effort required to structure communication and offers a more efficient approach to managing faculty work-related anxiety. These contributions are summarized in Table 1. Hard contributions and scope boundaries. Beyond domain adaptation of email task mining, Easy Assistant contributes a reliability-controlled and CLT-guided task-surfacing workflow that restructures unstructured faculty communications into prioritized task cards directly within the communication stream, rather than relying on post-hoc coping activities. Concretely, we make four technical and design contributions that are directly evaluated in this paper:

- C1. Selective task surfacing with reliability control. We operationalize Task Completeness (TC) as a weighted coverage-and-calibrated-confidence score and surface tasks only when TC $\geq$ 0.8, otherwise requesting one-tap human correction; we calibrate probabilities (temperature/Platt) and report reliability using ECE with 10 confidence bins.
- C2. Testable CLT-to-UI mapping rules. We translate CLT levers into concrete interaction constraints and implement them in focused task presentation (e.g., a single-screen "Today's Focus" surface gated by TC), information-stream collapsing to reduce noise, and a lightweight decomposition workflow that reduces plan-search costs.
- C3. Faculty-communication schema and reproducibility artifacts. We define an intent/slot task schema for faculty emails/chats/LMS notices, construct a de-identified dataset with stratified splits, and provide de-identified reproducibility resources (label schema, annotation guideline, preprocessing scripts, and configuration/evaluation code) listed in Supplementary Material.
- C4. Mechanism-supporting evidence (exploratory). In a 4-week single-group feasibility deployment (N=30), we observe pre–post reductions in GAD-7 and NASA-TLX alongside good usability, and we store minimal usage telemetry to enable future dose–response analyses linking exposure to TC-qualified tasks with outcome changes while avoiding causal over-claims.

We further clarify boundaries and differentiate our approach from prior task-mining systems through a structured comparison table in the Related Work section.

Table 1. Comparison between representative existing approaches and the Easy Assistant system

Innovation dimension	Existing research	This study
Theoretical framework	Most studies focus on traditional models of work stress	Integrates Cognitive Load Theory (CLT), the Job Demands–Resources (JD–R) model, and emotion-labor research to conceptualize unstructured digital messages as a specific job demand for university teachers.
Method innovation	Traditional task management systems focus on task sorting or classification	Implements an end-to-end NLP pipeline and a CLT-guided interface that transform unstructured messages into prioritized tasks with explicit uncertainty handling, thereby reducing extraneous cognitive load.

Application scenario innovation	A general work stress management tool aimed at employees or a wide range of people	Targets university teachers' day-to-day email and messaging overload rather than generic employee stress, embedding the system into their existing digital communication routines.

2. Related Work and Theoretical Background

2.1 Digital stress / mental-health interventions and engagement

Digital stress-management and mental-health interventions show small-to-moderate benefits across symptoms and settings, but reviews consistently note that engagement is fragile and depends on usability and fit with daily routines. We therefore target an upstream workplace stressor—message overload and triage—by restructuring fragmented communication into a smaller set of high-completeness tasks under reliability control. [1-6][8]

Collectively, this line of work motivates designing interventions that minimize friction and fit daily routines—an emphasis that informs our decision to target an upstream stressor in information work (turning fragmented communication into actionable plans) rather than prescribing additional coping exercises.

2.2 Inbox triage and task-centric interfaces (HCI)

HCI research has long explored inbox triage and task-centric interfaces as approaches to reduce user burden under heavy message streams. Early systems such as Taskmaster integrated task management directly into email, demonstrating the value of reframing email around actionable tasks rather than message-by-message processing [9]. Later work examined email triage behaviors and articulated design opportunities for next-generation clients, emphasizing rapid decision-making, prioritization support, and minimizing user effort during triage [10].

These task-centric foundations motivate our interface goals, while our system extends them by adding reliability-controlled NLP task surfacing and CLT-guided interaction constraints (summarized later in Table 2).

2.3 Uncertainty-aware NLP, selective prediction, and mobile deployment

In workplace communication, email triage imposes cognitive costs, and prior HCI systems motivate converting messages into tasks rather than forcing users to scan long inbox streams. Building on task-centric email and task extraction work, we use NLP to infer intents and slots and then apply TC-based gating to reduce exposure to low-confidence outputs. [11-16]

2.4 Evidence on effectiveness, engagement, and adoption of digital interventions

Beyond task extraction, we draw on selective prediction and human-in-the-loop ML: when confidence is low, systems should abstain or route outputs for user oversight. We instantiate this idea with calibrated probabilities and a conservative TC gate ($\tau = 0.8$), so that surfaced items satisfy explicit completeness-and-confidence criteria. [17-20]

Evidence from digital behavior-change and productivity interventions suggests that reducing friction and aligning tools with routines can improve engagement, but effects vary by context and study quality. Our design is positioned as a feasibility step, combining CLT-motivated interaction constraints with reliability control to support directional within-subject improvements without implying causality. [21-28]

In education-adjacent contexts, web-based stress-management interventions for beginning teachers and single-case or pragmatic designs in workplace settings provide complementary evidence on mechanisms of change and real-world feasibility [25, 26]. Feasibility studies of online mindfulness programs for teachers also suggest that educators are willing to engage with remote support when it fits around teaching and preparation [27]. Evidence syntheses of mobile mindfulness interventions further summarize psychological outcome gains while highlighting heterogeneity across designs [28].

Cognitive Load Theory (CLT) provides a mechanism: fragmented, multi-channel communication increases extraneous load (e.g., context switching and triage), while structuring information into goal-directed, task-sized units can free working memory for execution. We operationalize CLT through minimalism, segmentation, and short decision loops, and we evaluate whether this workflow is consistent with reduced workload and anxiety in a single-group feasibility study. [29-38]

2.5 Email overload, information overload, and NLP-based task extraction from emails/messages

Faculty work is heavily mediated by email and messaging, and work-email has been repeatedly linked to stress and difficulty detaching from work. A systematic review spanning 25 years of work-email research associates high email volume, frequent checking, and blurred temporal boundaries with elevated stress and impaired recovery [39]. More recent analyses show that specific email classes and work stressors predict higher perceived email load and poorer well-being [40]. These observations align with broader information-overload research showing that fragmented, poorly structured information can undermine performance and mental health [41].

The Job Demands–Resources (JD-R) framework provides a complementary perspective by conceptualizing burnout and engagement as outcomes of the balance between job demands and job resources, motivating interventions that add enabling resources to counter high demands [42].

On the NLP side, transformer-based models enable systems that mine intents and predict activities from incoming emails, supporting task identification and downstream organization [43]. Applied systems have also explored automated task extraction for workplace assistance in accessibility-oriented settings, where improving task clarity can be especially valuable [44].

Table 2 summarizes how Easy Assistant is positioned relative to these strands—task-centric HCI, calibrated / interactive uncertainty management, and email task mining—highlighting that our contribution is an end-to-end workflow that couples extraction with reliability-controlled task surfacing and CLT-guided interaction constraints.

Table 2. Comparison with prior inbox triage/task-centric interfaces, email task mining systems, and uncertainty-/deployment-aware NLP

Work	Domain	Output	Reliability handling	UI principle	Evaluation	Key delta vs. Easy Assistant
Bellotti et al. 2003 [9]	General email	Task-centric email tool (task management integrated into email)	Not explicit	Task-centric workflow	System design + user evaluation	We add NLP task extraction + calibrated selective surfacing (TC gate) + CLT-guided constraints and evaluate well-being/workload outcomes
Sarrafzadeh et al. 2019 [10]	General email	Empirical characterization of email triage (challenges/opportunities)	N/A (not an extraction model)	Triage behavior framing	User study (CHIIR)	We deliver a full pipeline (extract→gate→task cards) and evaluate in-situ feasibility for faculty workflows
Russell et al. 2024 [39]	Workplace email	Systematic review of work-email activity & outcomes	N/A	N/A	Systematic review	We move from “email-stress association” to an intervention that restructures communication into tasks
Khandaker et al. 2024 [43]	Incoming emails (general)	Intent/activity mining from emails (Transformers)	Not explicit	N/A	Offline modeling	We extend to end-user presentation with calibration + gating to control error exposure
Gollasch et al. 2025 [44]	Workplace (accessibility focus)	Automated task extraction from emails	Limited/unclear	Assistive extraction	Applied HCI venue	We add explicit uncertainty management (calibration + gate) + lightweight correction and link to faculty workload/anxiety
Guo et al. 2017 [11]	General ML	Calibration (temperature scaling)	Calibration metrics (ECE)	N/A	Benchmark study	We apply calibration in a task surfacing system and evaluate threshold trade-offs in context
Mosqueira-Rey et al. 2023 [12]	General ML	HITL survey	HITL principles	N/A	Survey	We operationalize HITL as one-tap correction tied to TC-gated surfacing and analyze failure modes
Xin et al. 2021 [13]	NLP	Selective prediction (abstention)	Abstention/selection under uncertainty	N/A	Offline experiments	Our TC gate is a system-level abstention mechanism aligned with usability constraints
Sanh et al. 2019 [16]	NLP	DistilBERT compression	Deployment efficiency	N/A	Model paper	Supports the need for efficient inference; ours focuses on workflow reliability + outcomes
Sun et al. 2020 [14]	NLP	MobileBERT	Deployment efficiency (resource-limited)	N/A	Model paper	Motivates mobile-ready NLP; we add reliability-controlled task surfacing
Jiao et al. 2020 [15]	NLP	TinyBERT	Deployment efficiency (distillation)	N/A	Model paper	Compression baseline; ours couples extraction + calibrated gating + CLT UI
Easy Assistant (ours)	Faculty communication	Structured task cards (intent + slots) and “Today’s Focus” surfacing	Calibration + TC gate + one-tap correction	CLT-guided interaction constraints	Offline evaluation + 4-week	—

feasibility
w/ scales

2. Methods

3.1 Research Design and Hypotheses

We adopted a two-stage mixed-methods approach to evaluate Easy Assistant. Stage I (Technical Evaluation) assessed the NLP pipeline's performance in converting unstructured text to tasks, using a labeled academic communication dataset and standard NLP metrics. Stage II (Efficacy Study) examined the system's impact on faculty's cognitive load and work anxiety via a single-group pretest-posttest design.

Three hypotheses were proposed:

- H1 (Technical Performance): The NLP pipeline will achieve F1-scores ≥0.90 for intent recognition and entity extraction, based on Transformer models' performance and domain-specific fine-tuning.
- H2 (Cognitive Load Reduction): Using the CLT-based interface for 4 weeks will significantly reduce cognitive load (≥10-point reduction on NASA-TLX).
- H3 (Anxiety Alleviation): Easy Assistant will reduce work anxiety (≥2-point reduction on GAD-7) after 4 weeks of use.

3.2 Data Collection and Procedures

3.2.1 Stage I: Technical Evaluation Data

A faculty messaging dataset (MDS) was created from institutional email, chat, and LMS notifications. Messages were de-identified and labeled for intent (TASK/MEETING/NOTICE/SOCIAL) and slots (e.g., deadline, location). Sampling was stratified by communication source and topic (teaching, research, administration), ensuring a representative mix of faculty communication.

3.2.2 Stage II: Efficacy Study Participants

Thirty faculty members (N=30) participated in a 4-week intervention. Participants were screened for work-related anxiety (GAD-7 ≥5), high digital communication usage (≥50 messages/day), and no prior Easy Assistant use.

Study Procedure:

- Pre-test (Week 0): Participants completed GAD-7 and NASA-TLX to establish baseline anxiety and cognitive load.
- Intervention (Weeks 1–4): Participants used Easy Assistant for task management (≥4 hours/day).
-
- Post-test (Week 5): Participants retook GAD-7 and NASA-TLX, and completed a post-intervention questionnaire.
- Follow-up (Week 8): Participants were invited to complete a brief check-in to assess whether any benefits persisted, but these contacts were purely qualitative and not included in the statistical analyses.

3.3 Measurement Instruments

Measurement instruments. Work anxiety was measured using the GAD-7 (total score 0–21; past-two-week symptoms). Subjective workload was measured using NASA-TLX (overall score 0–100) and we additionally analyze the Mental Demand subscale as a proxy for extraneous cognitive load. Measures were collected at Week 0 and Week 5 (pre–post); the Week-8 follow-up was qualitative only.

3.4 Metrics and decision rules

3.4.1 Task Completeness (TC)

Task Completeness (TC) used for gating and reporting is computed as Eq. (2).

Let the set of required slots be:

$$S=\{intent, deadline, recipient, deliverable, context\} \quad (1)$$

Each slot $s \in S$ has a non-negative weight w_s with $\sum_{s \in S} w_s = 1$.Denote by $1[s] \in \{0,1\}$ whether slot s is extracted and by $p_s \in [0,1]$ the calibrated probability that the extracted value is correct. We define:

$$\mathrm{TC}=\sum_{s \in S} w_s 1[s] p_s. \quad (2)$$

By construction, TC∈[0,1]; higher values indicate better coverage and confidence.Tasks are shown to users only if TC ≥ τ (with τ = 0.8 in our experiments); otherwise a one-tap human correction is requested.

3.4.2 Probability calibration and ECE

Classifier / posterior probabilities are calibrated (temperature / Platt as appropriate) on the validation set. Reliability is summarized by Expected Calibration Error (ECE): partition predictions into B bins of confidence; then:

$$\mathrm{ECE}=\sum_{b=1}^{B} \frac{n_b}{n} |\mathrm{acc}(b)-\mathrm{conf}(b)| \quad (3)$$

where n_b is the number of items in bin b, n is the total, and acc(b) and conf(b) are empirical accuracy and mean confidence of bin b, respectively.We use B = 10 equal-width confidence bins in all experiments.

By construction, ECE∈[0,1], and lower values indicate better calibration.

3.4.3 Priority score (P) and uncertainty

$$P = \alpha\cdot urgency+\beta\cdot intent\ risk+\gamma\cdot recipient\ importance +\delta\cdot(1\text{-}uncertainty) \quad (4)$$

With α, β, γ, δ ≥ 0 and α + β + γ + δ = 1. All four components (urgency, intent_risk, recipient_importance, 1 − uncertainty) are normalized to lie in [0,1], so P∈[0,1]. In our prototype deployment, the weights (α, β, γ, δ) are chosen once using an analytic hierarchy process (AHP): two domain experts provide pairwise preferences over the four criteria, and the resulting normalized eigenvector is used as (α, β, γ, δ). The same set of weights is kept fixed for all experiments; we do not tune them on the outcome metrics reported in this paper.

AHP criteria operationalization. The four components correspond to domain-relevant criteria used consistently across experiments: task urgency (deadline proximity), task importance (role/recipient criticality), task specificity (presence of actionable intent and required arguments), and uncertainty (calibrated model confidence complement). We fixed the resulting weights a priori and did not tune them on the outcome measures reported in this paper, ensuring that Stage II results are not optimized through post-hoc weight adjustment.

Here uncertainty is 1 − \bar{p}_x, the complement of the mean calibrated probability across required slots present. Priorities are mapped to {High, Medium, Low} using two thresholds $k_1 > k_2$ chosen on a held-out validation set to balance recall of important messages against alert fatigue (High if $P \geq k_1$; Medium if $k_2 \leq P < k_1$; otherwise Low). In practice, any thresholds with k_1 in the range [0.6, 0.8] and k_2 in the range [0.3, 0.5] led to very similar behavior, and we fix one such pair throughout all experiments.

3.4.4 Decision rule and tie-breaking

Tasks are first filtered using the TC gate (Eq. 2), then ranked by priority (Eq. 4), with ties broken by deadline and recipient importance.

Notes. Weights $\{w_s\}$ are fixed a priori based on domain heuristics and kept constant across experiments; all probabilities reported in Results are calibrated per §3.4.2.

3.5 Implementation Details and Reproducibility

We evaluated Easy Assistant on the de-identified faculty messaging dataset (MDS), spanning email, chat, and LMS sources. All personally identifying information was removed prior to labeling. Messages were annotated under a unified intent-and-slot schema to support downstream structuring into task cards (intent classification and entity/slot extraction). De-identification and labeling resources are provided in the Supplementary Material.

Annotation quality control. Two annotators labeled messages independently; disagreements were adjudicated. The full schema, guidelines, and decision rules are in the Supplementary Material.

3.6 Data protection and confidentiality

Data protection and confidentiality. Stage-I technical data were de-identified before labeling; normalization removed signatures, quoted text, and legal disclaimers. For Stage-II, only study IDs, GAD-7/NASA-TLX responses, and minimal usage telemetry were stored; no raw emails, chat logs, or LMS messages were persisted. Data in transit and at rest were encrypted; access to potentially sensitive artifacts followed role-based and attribute-based controls with audit logging.

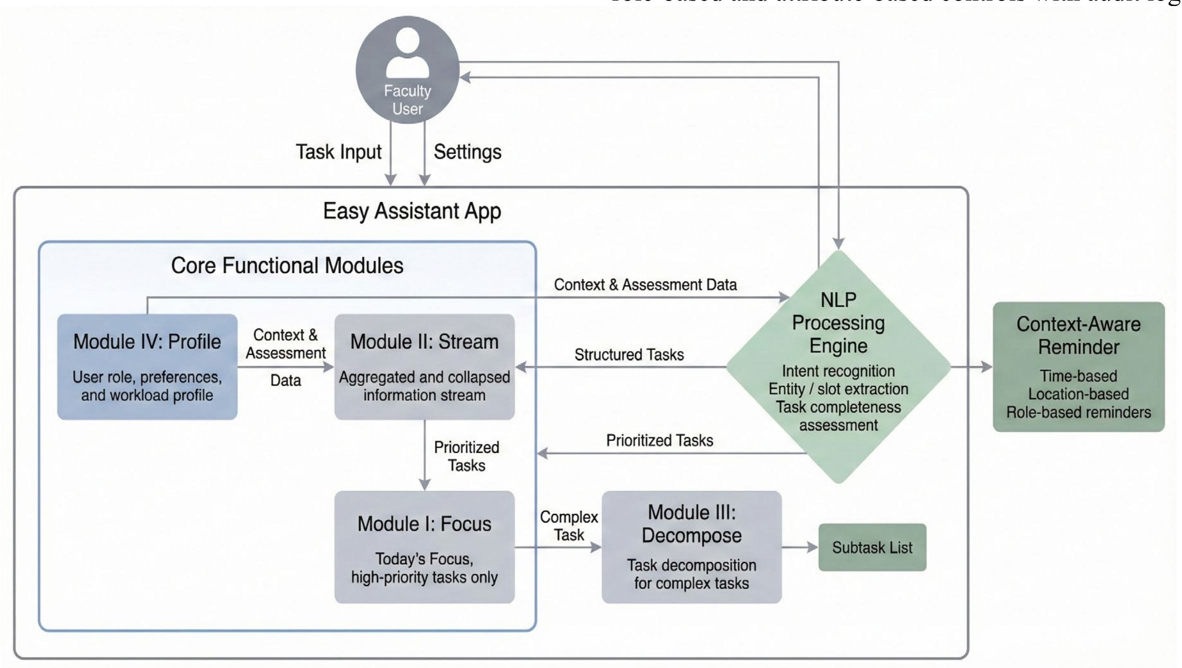


Figure 1. System overview and four-stage pipeline (cleaning → intent/entity recognition → structuring & prioritization). Calibrated confidences feed the task-completeness gate (Eq. (2); see 3.4.1), and reliability is quantified with ECE (Eq. (3); see 3.4.2).

3. System Design

4.1 System Architecture

System architecture. Easy Assistant follows a client–server design (Figure. 1). The mobile client collects faculty messages from email/chat/LMS sources, surfaces only TC-gated task cards, and supports lightweight correction. Server-side components normalize and secure inputs, run the four-stage NLP pipeline to produce structured Task objects with calibrated confidence, TC, and priority, and store tasks to support scheduling, decomposition, and reminders. Implementation and deployment specifics are documented in the Supplementary Material.

4.2 NLP Intelligent Information Processing Pipeline

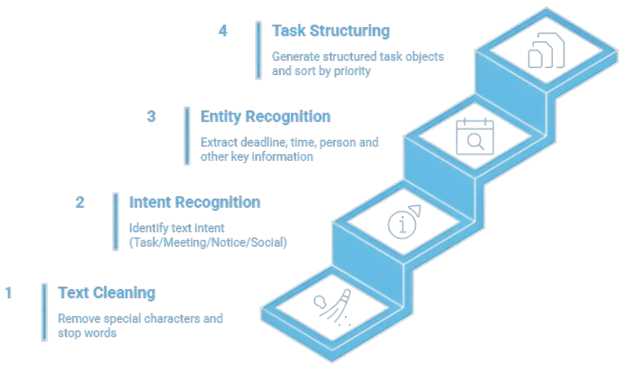


Figure 2. NLP Processing Flow

As shown in Figure 2, we implement a four-stage cascaded pipeline—Cleaning, Intent Recognition, Entity/Slot Recognition, and Structuring & Prioritization—to convert faculty messages into executable task cards. The pipeline is modular, while reliability-controlled surfacing is enforced by calibrated confidence and a Task Completeness (TC) acceptance gate.

4.2.1 Stage 1: Text Cleaning and Normalization

We remove channel-specific artifacts and normalize surface forms to reduce noise while preserving traceability. Cleaning includes signature and quoted/forwarded-text removal, URL and emoji normalization, whitespace/bullet normalization, duplicate-paragraph merging, and noisy-paragraph filtering (examples in Figure 3). Importantly, we maintain character-level span mappings between raw and cleaned text so extracted slots can be verified in the original message, supporting user trust.

4.2.2 Stage 2: Intent Recognition

The intent stage identifies the task type (or non-action messages) and selects the downstream slot schema. We formulate intent recognition as multi-label classification with a transformer encoder. For long messages, we apply segmented encoding and aggregate segment-level predictions via attention-weighted averaging. To reduce overconfidence, we calibrate output probabilities using temperature scaling, targeting ECE ≤ 0.03.

4.2.3 Stage 3: Entity/Slot Recognition

We extract key task arguments (e.g., action, recipient, deadline, and context) via sequence labeling using a Transformer–CRF model. Slot labels follow the unified intent-and-slot schema; training is strengthened with template-based augmentation and back-translation to improve robustness to phrasing variability and mixed-context messages.

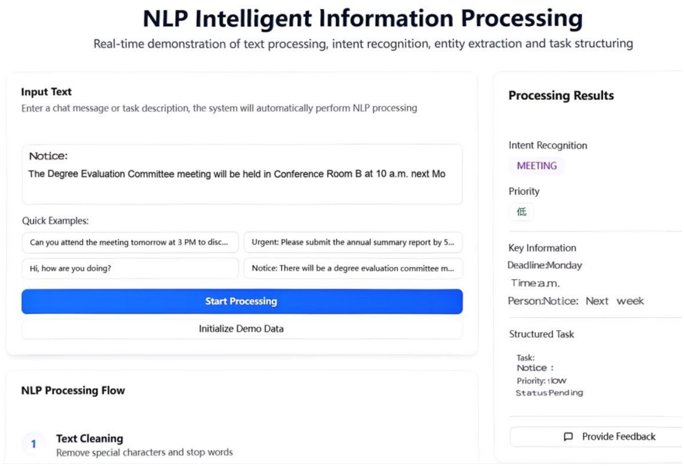


Figure 3. NLP Intelligent Information Processing Examples

4.2.4 Stage 4: Structuring and Prioritization

Upstream outputs are aggregated into a structured Task object. We compute uncertainty from calibrated confidences and define Task Completeness (TC) as the weighted coverage-and-confidence over required slots (Eq. (3)). Tasks are surfaced only if TC ≥ 0.8 (Eq. (2)). Surfaced tasks are then ranked by a priority score combining urgency, intent_risk, recipient_importance, and $(1 - \text{uncertainty})$ (Eq. (4)); ties are broken by earlier deadlines.

4.2.5 Uncertainty Propagation and Acceptance Gate

To suppress error propagation across stages, uncertainty is propagated through the pipeline and explicitly constrained by the TC acceptance gate. User corrections are incorporated via lightweight online updates so the system can gradually reduce recurring extraction errors while maintaining conservative surfacing behavior.

4.3 Interface and Interaction Design: CLT-Driven Principles

The interface is strictly constrained by Cognitive Load Theory (CLT) to reduce extraneous load, manage intrinsic load, and appropriately stimulate germane load. Table 3 summarizes how we operationalized five core CLT principles into concrete UI and interaction rules applied throughout the app.

Figure 4 shows the Today's Focus screen, whose primary surface presents only task cards that meet the acceptance threshold and have TC $\geq$ 0.8 (see Eq. 2), enabling single-screen focus and minimizing context switching.

Table 3. Mapping of Cognitive Load Theory (CLT) Design Principles to UI/Interaction Correspondences

Design principles (CLT)	UI/interaction correspondences
Minimalist visuals; eliminate noise	Low-density layout, ample white space, low-saturation palette; secondary streams collapsed by default
Focus on primary tasks (goal-directedness)	"Today's Focus" surface shows only high-priority, actionable task cards
Segmentation (step-by-step guidance)	"Decomposition Guide" breaks complex tasks into sequential, executable steps
Modality/contiguity	Contextual nudges/reminders by time, location, and role; evidence localization for quick verification
Germane load (feedback & reinforcement)	Completion cues and lightweight progress visualization to support schema consolidation

Notes. CLT = Cognitive Load Theory. Principles mapped from Section 4.3 design rules used in the app.

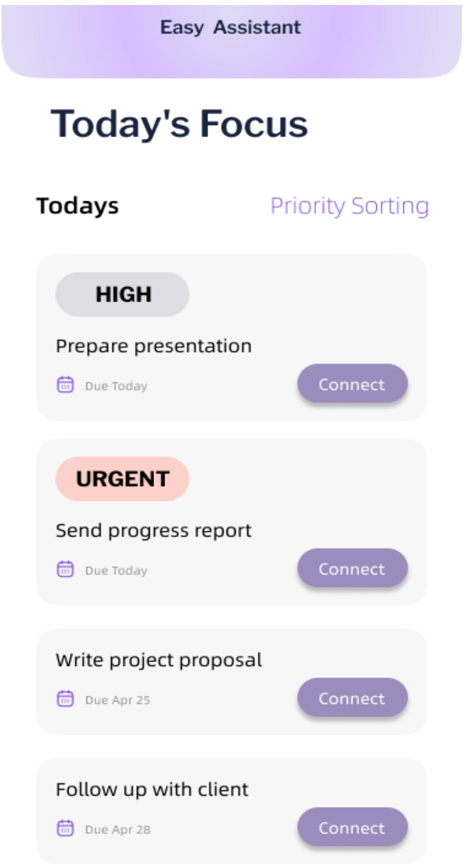


Figure 4. "Today's Focus" screen showing priority-ordered task cards gated by TC ≥ 0.8; essential fields (due time, counterpart, source tags, brief description); one-tap actions; and evidence localization. Notes. TC = Task Completeness; tasks are surfaced only if TC ≥ 0.8.

Information Stream Collapse. The "Stream" feed aggregates all incoming messages with intelligent collapsing: social and notice items are folded by default to reduce visual noise; users can expand any item for verification without interrupting the main flow; and correction operations are coupled to model uncertainty, creating a low-cost human-in-the-loop cycle.

Figure 5 illustrates the Decomposition Guide, which provides a two-step workflow for breaking down complex tasks: (1) identifying major milestones and (2) generating the first executable actions for each milestone. The system writes the resulting subtasks back to the parent card, reducing plan-search costs and supporting intrinsic-load management.

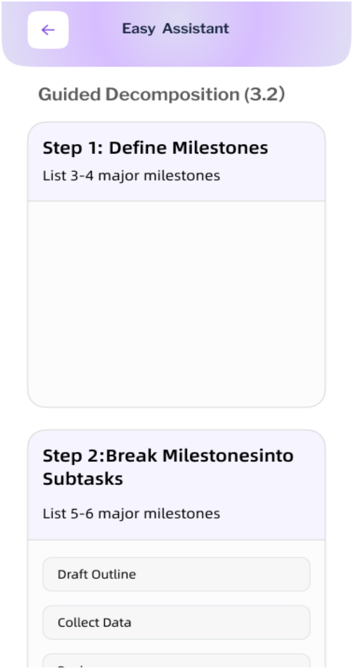


Figure 5. “Decomposition Guide” for complex tasks: milestone identification and first executable action generation; subtasks are written back to the parent card to reduce plan-search costs. Contextual Reminders. Lightweight nudges are triggered by context: time (e.g., “You have a task due today”), location (e.g., “You’re in the lab; check the experiment”), role/activity (e.g., “You’re in a meeting; discuss the timeline”), and device state (e.g., “You’re not in a meeting; now is a good time to work on that report”).

Figure 6. Wellness dashboard showing task-completion trends, GAD-7 trajectories, and NASA-TLX workload trends, providing lightweight progress visualization.

Figure 6 presents the Wellness Dashboard, which provides feedback on task completion, anxiety, and cognitive load, including task-completion trends, GAD-7 trajectories, NASA-TLX trends, and lightweight progress visualizations that inform without overwhelming.

4. Evaluation and Results

5.1 Technical Evaluation Results (Stage I)

Technical Evaluation Results (Stage I). The pipeline achieved strong performance, with macro-averaged intent F1 = 0.92 and macro-averaged entity/slot F1 = 0.89 (Table 4). For reliability-controlled surfacing, the TC gate pass rate at τ = 0.8 was 87%, and the mean TC over all extracted tasks was 0.83, thereby meeting the Stage I targets (Table 4). In the operational setting, Table 4b summarizes TC-gated surfacing at τ = 0.8, where 87% of extracted tasks are surfaced (proxy correction burden = 13%).

Structured error analysis. We observed three recurring error modes that directly affect TC-gated surfacing. First, slot omission occurs when key arguments (e.g., deadlines or recipients) are missed in long, multi-intent messages; this typically lowers TC and results in conservative non-surfacing under τ = 0.8. Second, boundary drift arises when the model assigns a slot span to adjacent context (e.g., quoting or forwarded text), producing partially correct but misaligned entities; calibration mitigates overconfident surfacing by reducing the likelihood that such outputs pass the gate. Third, implicit-task phrasing (e.g., socially embedded requests without explicit action verbs) can be correctly classified at the intent level but remain underspecified for slot filling, again lowering TC. These patterns suggest that TC gating primarily reduces user exposure to under-specified or mislocalized tasks, at the cost of withholding some borderline tasks that may require human clarification.

Deployment feasibility metrics on mobile devices (latency, throughput, energy, and crash-free rate) met the deployment targets; detailed results are reported in the Supplementary Material.

Table 4a. NLP Pipeline Performance Metrics

Metric	Intent Recognition	Entity/Slot Extraction
Macro-averaged F1	0.92	0.89
Micro-averaged F1	0.94	0.91
Precision	0.93	0.90
Recall	0.91	0.88

Note. Reliability-controlled surfacing: the TC gate pass rate at $\tau = 0.8$ is 87%, and the mean TC over all extracted tasks is 0.83.

Table 4b. Operational TC-gated task surfacing summary ($\tau = 0.8$).

τ(operational)	Tasks surfaced (%)	Proxy correction burden (%, 1 − surfaced)
0.8	87	13

Note. Proxy correction burden is computed as 1 − (tasks surfaced).

5.2 Efficacy Study Results (Stage II – feasibility stage)

Participant characteristics. Thirty faculty participants (15 male, 15 female) completed the feasibility study. Participants had a mean age of 42.3 years (SD = 8.5) and a mean of 12.4 years of academic experience (SD = 7.2). Disciplines included Engineering (n = 10), Business (n = 8), Science (n = 7), and Humanities (n = 5). At baseline, participants reported moderate work anxiety on average (GAD-7 mean = 10.5, SD = 2.1) and high message volume (mean daily emails/messages = 87.3, SD = 23.4).

Sample flow and missing data. All 30 enrolled participants completed both baseline (Week 0) and post-intervention (Week 5) assessments for GAD-7 and NASA-TLX. Analyses were conducted on a complete-case basis with 30 paired observations for each outcome. The recruitment and retention process is summarized in the participant flow diagram (Supplementary Figure S1).

Primary outcomes (pre–post). Over the feasibility period (4-week intervention with post-assessment at Week 5), participants showed statistically significant pre–post reductions in both anxiety and subjective workload (Table 5). GAD-7 decreased from 10.5 (SD = 2.1) to 7.8 (SD = 1.9), with a mean reduction (Pre − Post) of 2.7 points (95% CI [1.8, 3.6]; $t = 5.42$; $d = 0.99$). Overall NASA-TLX decreased from 65.2 (SD = 8.5) to 48.9 (SD = 7.1), with a mean reduction of 16.3 points (95% CI [12.1, 20.5]; $t = 7.91$; $d = 1.44$). Mental Demand decreased from 72.1 (SD = 9.3) to 52.4 (SD = 8.2), with a mean reduction of 19.7 points (95% CI [15.2, 24.2]; $t = 8.15$; $d = 1.49$). Interpretation in relation to the proposed mechanism. The observed reductions in workload—especially Mental Demand (mean reduction = 19.7 points; $d = 1.49$)—are consistent with the proposed CLT mechanism in which restructuring fragmented communication into a smaller set of high-completeness tasks reduces extraneous cognitive load during triage. Given the single-group feasibility design, these results are interpreted as within-subject changes consistent with the hypothesized direction of improvement, without implying causal effects.

Engagement and task completion. Participants completed an average of 78.3% of tasks (SD = 12.1%), and 92% of participants completed at least 70% of their tasks during the study period. These completion patterns are consistent with sustained engagement during the feasibility stage.

User satisfaction. Post-intervention usability ratings indicated good usability, with a mean System Usability Scale (SUS) score of 78.2 (SD = 10.3).

Table 5. Pre–post changes in GAD-7 and NASA-TLX (paired-samples t-tests; n = 30).

Outcome	Pre-test Mean (SD)	Post-test Mean (SD)	Mean Difference (95% CI)	t-statistic	p-value	Cohen's d
GAD-7 (Work Anxiety)	10.5 (2.1)	7.8 (1.9)	2.7 (1.8–3.6)	5.42	p < 0.01	0.99
NASA-TLX (Cognitive Load)	65.2 (8.5)	48.9 (7.1)	16.3 (12.1–20.5)	7.91	p < 0.001	1.44
NASA-TLX Mental Demand	72.1 (9.3)	52.4 (8.2)	19.7 (15.2–24.2)	8.15	p < 0.001	1.49

Notes.Values are means (SD). Mean differences are computed as Pre − Post; positive values indicate reductions from baseline. Effect sizes are Cohen's d for within-subject designs.

Table 6. Hypotheses verification summary

ID	Hypothesis	Decision rule (analysis)	Evidence (this study)	Verdict
H1	The NLP pipeline meets the pre-specified (not formally registered) technical targets.	Intent/Entity F1, mean TC, and proportion of tasks with TC ≥ 0.8 (descriptive metrics; Stage I).	Intent F1 = 0.92; Entity F1 = 0.89; Mean TC = 0.83; 87% of tasks with TC ≥ 0.8.	Consistent with expectations in this feasibility stage
H2	Participants show pre–post reductions in subjective workload.	Paired-samples t-test on NASA-TLX (overall), two-tailed α = 0.05.	Δ = −16.3, 95% CI [−20.5, −12.1]; t(29) = 7.91, p < 0.001, Cohen's d = 1.44.	Consistent with expectations in this feasibility stage
H3	Participants show pre–post reductions in work anxiety.	Paired-samples t-test on GAD-7, two-tailed α = 0.05.	Δ = −2.7, 95% CI [−3.6, −1.8]; t(29) = 5.42, p < 0.01, Cohen's d = 0.99.	Consistent with expectations in this feasibility stage

Notes. Negative Δ denotes pre–post reductions; tests are two-tailed. TC = Task Completeness; the acceptance gate is TC ≥ 0.8 (see Eq. (1)).

5.3 Hypotheses Verification Table

Table 6 summarizes the verification results for the pre-specified (not formally registered) hypotheses across the two study stages. H1 concerns the technical performance of the NLP pipeline (Stage I), whereas H2–H3 concern psychological outcomes (Stage II). Results were consistent with all three hypotheses in this feasibility stage.

5. Discussion

This work provides feasibility evidence for a CLT-guided, NLP-enabled workflow that structures faculty communication into actionable tasks while controlling reliability. In Stage I, the pipeline achieved strong extraction performance (intent F1 = 0.92; entity/slot F1 = 0.89) and conservative, reliability-controlled surfacing at the operational threshold $\tau = 0.8$, where 87% of extracted tasks passed the TC gate and mean TC over all extracted tasks was 0.83 (Tables 4 and 4b). These results indicate that task surfacing can be coupled with explicit acceptance criteria to reduce exposure to low-confidence outputs.

Beyond the main metrics, the offline ablation results help clarify which pipeline components contribute most to end-to-end quality. This pattern supports the design choice of treating the pipeline as an integrated system rather than relying on any single model component.

In Stage II (n = 30), participants showed statistically significant pre–post reductions in both work anxiety and subjective workload (Table 5). GAD-7 decreased by 2.7 points (95% CI [1.8, 3.6]; d = 0.99), and overall NASA-TLX decreased by 16.3 points (95% CI [12.1, 20.5]; d = 1.44). Mental Demand decreased by 19.7 points (95% CI [15.2, 24.2]; d = 1.49). These changes are visualized in Supplementary Figure S2 (left), which summarizes pre–post outcome means with 95% confidence intervals. Engagement and usability were also encouraging (mean task completion = 78.3%, SD = 12.1%; 92% completing ≥70%; SUS = 78.2, SD = 10.3), suggesting that the system can be integrated into faculty routines with acceptable interaction overhead.

A key contribution is the way CLT is operationalized as concrete interaction constraints aligned with reliability control. Faculty communication imposes extraneous load when information is fragmented across channels, mixed with irrelevant content, and requires repeated context switching. Easy Assistant addresses this by emphasizing minimalism (surface fewer, higher-quality items), segmentation (task-sized units rather than long message threads), goal-directed focus (actionable tasks instead of inbox streams), and short decision loops (quick actions and progressive disclosure). The TC gate complements these constraints by making surfacing conservative: rather than maximizing coverage, the assistant prioritizes reliability and reduces the likelihood that users must spend effort verifying uncertain outputs. The reductions in overall NASA-TLX and especially Mental Demand are consistent with this mechanism, suggesting that restructuring communication into clearer tasks may help users allocate cognitive resources to execution rather than triage. At the same time, given the single-group feasibility design, these findings should be interpreted as within-subject changes consistent with the hypothesized direction of improvement, without implying causal effects.

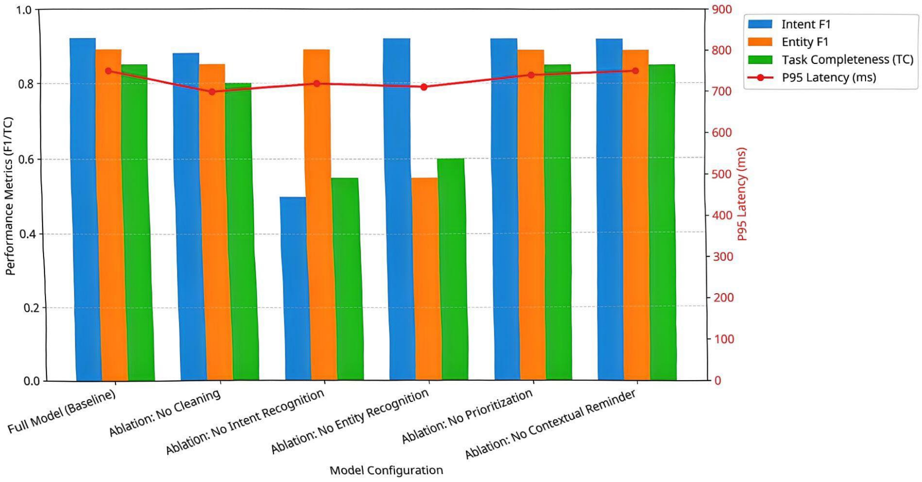
Finally, mobile feasibility is supported by the deployment trend shown in Supplementary Figure S3 (right), which summarizes P95 latency (ms) during deployment testing by week. Together with the extraction and outcome results, these figures strengthen the evidence chain from technical performance to practical deployability.

Given the single-group feasibility design, we interpret the results as directional within-subject changes that are consistent with the hypothesized mechanism, without implying causal effects.

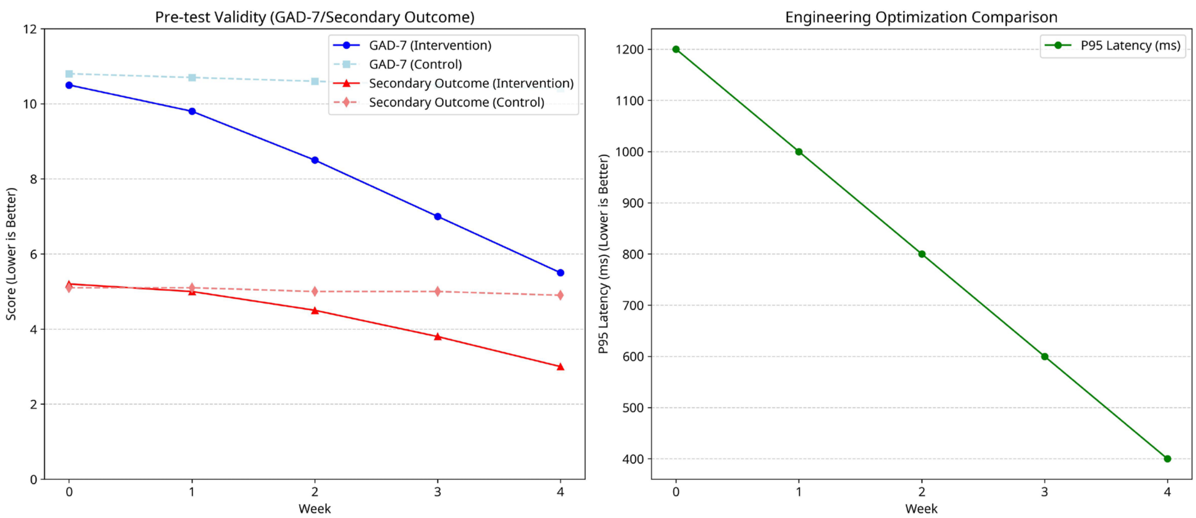
6.1 Limitations and Future Work

This study has important limitations. First, Stage II used a single-group pre–post feasibility design (n = 30). Without a control condition, observed changes may reflect time trends, regression to the mean, expectancy effects, or concurrent workload changes. Second, outcomes were self-reported (GAD-7 and NASA-TLX). Future studies should triangulate subjective measures with behavioral metrics such as after-hours message handling, time-to-task completion, or interaction logs, and should evaluate longer-term persistence beyond the post-intervention timepoint.In addition, we did not conduct a full threshold-sensitivity study across multiple τ values within the main paper, and different users may prefer different coverage–reliability trade-offs. We also did not model dose–response relationships between usage intensity and outcome changes in this feasibility sample; future controlled studies with richer usage logging will be required to test mechanism-linked associations.

On the technical side, performance estimates were obtained on a de-identified faculty messaging dataset; generalization across institutions, disciplines, or communication cultures may require additional adaptation. Although conservative gating improves reliability, it can also withhold some potentially useful tasks; different users may prefer different trade-offs between coverage and error exposure. Future work will (i) evaluate efficacy in controlled trials (e.g., randomized or stepped-wedge designs) with longer deployments, (ii) personalize TC weighting and the gate threshold based on user tolerance for correction effort, and (iii) conduct ablations that explicitly link component-level changes to both offline metrics and deployment constraints, building on the analyses already summarized in Supplementary Material (Supplementary Figure S2) and the deployment trajectory in Supplementary Figure S3.



Supplementary Figure S2. Offline Technical Indicators and Ablation Results.



Supplementary Figure S3. Pre–post outcomes and engineering latency. Left: GAD-7 means at Week 0 and Week 5 with 95% confidence intervals. Right: P95 latency (ms) by week during deployment testing.

6. Conclusion

This paper presented Easy Assistant, a CLT-guided and reliability-controlled task surfacing system that converts fragmented faculty communications into actionable tasks. Technically, the NLP pipeline achieved strong extraction performance (intent F1 = 0.92; entity/slot F1 = 0.89) and conservative operational surfacing at $\tau = 0.8$, with 87% of extracted tasks passing the TC gate and mean TC = 0.83 (Tables 4 and 4b). Offline ablations further highlight the contributions of key pipeline stages and the feasibility of engineering optimizations (Supplementary Material, Supplementary Figure S1).

In a single-group feasibility study (n = 30), participants showed statistically significant pre–post reductions in both work anxiety and subjective workload (Table 5; Supplementary Figure S3), alongside sustained engagement (mean task completion = 78.3%; 92% completing ≥70%) and good usability (SUS = 78.2). While controlled trials are required to establish causal effects, these results support the practicality of combining uncertainty-aware NLP with CLT-informed interaction constraints to reduce perceived cognitive burden in everyday information work. Future work will evaluate efficacy under controlled designs, explore personalization of TC weighting and gate thresholds, and validate robustness across broader institutional settings.

Conflict of Interest Statement

No potential conflict of interest was reported by the authors.

Ethical Statement

The study was approved by the Institutional Review Board of Zhongkai University of Agriculture and Engineering. Informed consent was obtained from all participants.

Acknowledgements.

This research was funded by the 2024 Guangdong Provincial Department of Science and Technology "Guangdong Rural Science and Technology Commissioner Project: Jiangdong Vegetable, Rice, and Agro-Cultural Tourism Industry Brand Building and Promotion" (Project No. KTP20240829); the Construction of science and technology plan project content in Jiangdong Town, Chaoan District, Chaozhou City (Project No. D125211E7); and 2024 Chaozhou City Chaoan District Science and Technology Special Fund: Construction of Ecological Industry Innovation Talent Cultivation Base under the "Hundred, Thousand, Ten Thousand Project" (Specialized Town Transformation and Upgrading Project).

Contribution statement

Jiao Jiang (Conceptualization, Methodology, Investigation, Visualization, Writing – original draft, Project administration); Haoteng Chen (Software, Data curation, Formal analysis, Validation, Visualization, Writing – review & editing); Hao Gu (Supervision, Project administration, Funding acquisition, Writing – review & editing); Junfeng Chen (Conceptualization, Methodology, Supervision, Funding acquisition, Writing – review & editing); Yesheng Hong (Software, Data curation, Formal analysis, Resources, Writing – review & editing). All authors read and approved the final manuscript.

References

[1] Manning JB, Blandford A, Edbrooke-Childs J. Digital companion choice to support teachers' stress self-management: systematic approach through taxonomy creation. JMIR Form Res. 2022;6(2):e32312. doi:10.2196/32312.

[2] Dantes GR, et al. Brief mobile app–based mindfulness intervention for Indonesian senior high school teachers: protocol for a pilot randomized controlled trial. JMIR Res Protoc. 2024;13(1):e56693. doi:10.2196/56693.

[3] Callahan C, Kimber J, Hu E, Tanner L, Kunkle S. The real-world impact of app-based mindfulness on Headspace members with moderate and severe perceived stress: observational study. JMIR Mhealth Uhealth. 2024;12(1):e52968. doi:10.2196/52968.

[4] McCloud T, Jones R, Lewis G, Bell V, Tsakanikos E. Effectiveness of a mobile app intervention for anxiety and depression symptoms in university students: randomized controlled trial. J Med Internet Res. 2020;22(7):e15418. doi:10.2196/15418.

[5] Cameron G, et al. Effectiveness of digital mental health interventions in the workplace: umbrella review of systematic reviews. JMIR Ment Health. 2025;12:e67785. doi:10.2196/67785.

[6] Sîrbu V, David OA. Efficacy of app-based mobile health interventions for stress management: a systematic review and meta-analysis. Clin Psychol Rev. 2024;114:102515. doi:10.1016/j.cpr.2024.102515.

[7] C. He and Y. Deng, "05 AI-assisted inclusive interaction design system for users with cognitive differences", artsci, vol. 25, no. 1, pp. 1–17, Jul. 2025.

[8] Wang H. Research on Technostress in Human-Computer Interaction. Philosopher's Compass. 2024 Oct 1;1(1):43-50.

[9] Bellotti V, Ducheneaut N, Howard M, Smith I. Taking email to task: the design and evaluation of a task management centered email tool. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03); 2003; Ft. Lauderdale, FL, USA. p.345-352. doi:10.1145/642611.642672.

[10] Sarrafzadeh B, Awadallah AH, Shokouhi M. Exploring email triage: challenges and opportunities. In: Proceedings of the 2019 Conference on Human Information Interaction and Retrieval (CHIIR '19); 2019; Glasgow, Scotland, UK. p.325-329. doi:10.1145/3295750.3298960.

[11] Guo C, Pleiss G, Sun Y, Weinberger KQ. On calibration of modern neural networks. In: Proceedings of the 34th International Conference on Machine Learning (ICML 2017). Proc Mach Learn Res. 2017; 70:1321-1330. doi:10.48550/arXiv.1706.04599.

[12] Mosqueira-Rey E, Hernández-Pereira E, Alonso-Ríos D, Bobes-Bascarán J, Fernández-Leal Á. Human-in-the-loop machine learning: a state of the art. Artif Intell Rev. 2023;56(4):3005-3054. doi:10.1007/s10462-022-10246-w.

[13] Xin J, Tang R, Yu Y, Lin J. The Art of Abstention: Selective Prediction and Error Regularization for Natural Language Processing. In: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers); 2021; Online. p.1040–1051. doi:10.18653/v1/2021.acl-long.84.

[14] Sun Z, Yu H, Song X, Liu R, Yang Y, Zhou D. MobileBERT: a Compact Task-Agnostic BERT for Resource-Limited Devices. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics; 2020; Online. p.2158–2170. doi:10.18653/v1/2020.acl-main.195.

[15] Jiao X, Yin Y, Shang L, Jiang X, Chen X, Li L, et al. TinyBERT: Distilling BERT for Natural Language Understanding. In: Findings of the Association for Computational Linguistics: EMNLP 2020; 2020; Online. p.4163–4174. doi:10.18653/v1/2020.findings-emnlp.372.

[16] Sanh V, Debut L, Chaumond J, Wolf T. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. arXiv. 2019. doi:10.48550/arXiv.1910.01108.

[17] Hwang H, Kim SM, Netterstrøm B, Han DH. The efficacy of a smartphone-based app on stress reduction: randomized controlled trial. J Med Internet Res. 2022;24(2):e28703. doi:10.2196/28703.

[18] Moberg C, Niles A, Beermann D. Guided self-help works: randomized waitlist-controlled trial of Pacifica, a mobile app integrating cognitive behavioral therapy and mindfulness for stress, anxiety, and depression. J Med Internet Res. 2019;21(6):e12556. doi:10.2196/12556.

[19] Linardon J, et al. Current evidence on the efficacy of mental health smartphone apps for symptoms of depression and anxiety: a meta-analysis of 176 randomized controlled trials. World Psychiatry. 2024; 23(1): 139-149. doi:10.1002/wps.21183.

[20] Linardon J, Firth J, Torous J, Messer M, Fuller-Tyszkiewicz M. Efficacy of mental health smartphone apps on stress levels: a meta-analysis of randomized controlled trials. Health Psychol Rev. 2024; 18(4): 839-852. doi:10.1080/ 17437199.2024.2379784.

[21] Gan DZ, McGillivray L, Larsen ME, Christensen H, Torok M. Technology-supported strategies for promoting user

engagement with digital mental health interventions: a systematic review. Digit Health. 2022;8:20552076221098268. doi:10.1177/20552076221098268.
[22] Cheng VWS, Davenport T, Johnson D, Vella K, Hickie IB. Gamification in apps and technologies for improving mental health and well-being: systematic review. JMIR Ment Health. 2019;6(6):e13717. doi:10.2196/13717.
[23] Hwang WJ, Jo HH. Evaluation of the effectiveness of a mobile app-based stress-management program: a randomized controlled trial. Int J Environ Res Public Health. 2019;16(21):4270. doi:10.3390/ijerph16214270.
[24] Coelhoso CC, Tobo PR, Lacerda SS, Lima AH, Barrichello CRC, Amaro Jr E, et al. A new mental health mobile app for well-being and stress reduction in working women: randomized controlled trial. J Med Internet Res. 2019;21(11):e14269. doi:10.2196/14269.
[25] Heckendorf H, Lehr D. Efficacy of a web-based stress management intervention for beginning teachers on reducing stress and mechanisms of change: randomized controlled trial. J Med Internet Res. 2025;27:e58475. doi:10.2196/58475.
[26] Demirel S, et al. Assessing the effectiveness of STAPP@Work, a self-management mobile app, in reducing work stress and preventing burnout: single-case experimental design study. J Med Internet Res. 2024;26:e48883. doi:10.2196/48883.
[27] Mendelson T, Webb L, Artola A, Molinaro M, Sibinga E. An online mindfulness program for teachers: a feasibility study of the DeStress Monday at School program. Mindfulness. 2023;14(6):1419–1434. doi:10.1007/s12671-023-02142-3.
[28] Tan ZYA, Wong SH, Cheng LJ, Lau ST. Effectiveness of mobile-based mindfulness interventions in improving mindfulness skills and psychological outcomes for adults: a systematic review and meta-regression. Mindfulness. 2022;13(10):2379-2395. doi:10.1007/s12671-022-01962-z.
[29] East ML, Havard B, Hastings NB. Mental health mobile apps' instruction: technology adoption theories applied in a mixed methods study of counseling faculty. J Technol Hum Serv. 2016;34(4):301-325. doi:10.1080/15228835.2016. 1233842.
[30] Sawrikar V, Mote K. Technology acceptance and trust: overlooked considerations in young people's use of digital mental health interventions. Health Policy Technol. 2022;11(4):100686. doi:10.1016/j.hlpt.2022.100686.
[31] Uribe F A R, Favacho M F M, Moura P M N, et al. Effectiveness of an app-based intervention to improve well-being through cultivating positive thinking and positive emotions in an adult sample: study protocol for a randomized controlled trial. Front Psychol. 2023; 14: 1200960.
[32] Champion L, Economides M, Chandler C. The efficacy of a brief app-based mindfulness intervention on psychosocial outcomes in healthy adults: a pilot randomized controlled trial. PLoS One. 2018;13(12):e0209482. doi:10.1371/journal.pone.0209482.
[33] Shankland R, Rosset E. Review of brief school-based positive psychological interventions: a taster for teachers and educators. Educ Psychol Rev. 2017; 29(2): 363-392. doi:10.1007/s10648-016-9357-3.
[34] Weber S, Lorenz C, Hemmings N. Improving stress and positive mental health at work via an app-based intervention: a large-scale multi-center randomized control trial. Front Psychol. 2019;10:2745. doi:10.3389/fpsyg.2019.02745.
[35] Hendriks T, Schotanus-Dijkstra M, Hassankhan A, De Jong J, Bohlmeijer E. The efficacy of multi-component positive psychology interventions: a systematic review and meta-analysis of randomized controlled trials. J Happiness Stud. 2020;21(1):357-390. doi:10.1007/s10902-019-00082-1.
[36] Koh J, Tng GY, Hartanto A. Potential and pitfalls of mobile mental health apps in traditional treatment: an umbrella review. J Pers Med. 2022; 12(9): 1376. doi:10.3390/jpm12091376.
[37] Miguel C, et al. Universal, selective and indicated interventions for supporting mental health at the workplace: an umbrella review of meta-analyses. Occup Environ Med. 2023;80(4):225-236. doi:10.1136/oemed-2022-108698.
[38] Chong SOK, Pedron S, Abdelmalak N, Laxy M, Stephan AJ. An umbrella review of effectiveness and efficacy trials for app-based health interventions. NPJ Digit Med. 2023;6(1):233. doi:10.1038/s41746-023-00981-x.
[39] Russell E, Meyer B, de Jonge J, Sonnentag S, van den Heuvel M. Getting on top of work-email: A systematic review of 25 years of research on work-email activity. J Occup Organ Psychol. 2024;97:74–103. doi:10.1111/joop.12462.
[40] Kern P, Ohly S, Ďuranová L, Friedrichs J. Drowning in emails: investigating email classes and work stressors as antecedents of high email load and implications for well-being. Front Psychol. 2024; 15:1439070. doi:10.3389/fpsyg.2024.1439070.
[41] Arnold M, Goldschmitt M, Rigotti T. Dealing with information overload: a comprehensive review. Front Psychol. 2023;14:1122200. doi:10.3389/fpsyg.2023.1122200.
[42] Bakker AB, Demerouti E. Job demands–resources theory: taking stock and looking forward. J Occup Health Psychol. 2017;22(3):273–285. doi:10.1037/ocp0000056.
[43] Khandaker MRU, Alassafi MO, Islam MR, Chen L. Transformer models for mining intents and predicting activities from incoming emails. Eng Appl Artif Intell. 2024;127:107450. doi:10.1016/j.engappai.2023.107450.
[44] Gollasch D, Branig M, Dähne H, Weber G. Automated task extraction from emails: aiding autistic individuals in the workplace. In: Antona M, Stephanidis C, editors. Universal Access in Human–Computer Interaction. 19th International Conference, UAHCI 2025, Held as Part of the 27th HCI International Conference, HCII 2025, Gothenburg, Sweden, June 22–27, 2025. Proceedings, Part I. Cham: Springer; 2025. p.343–359.