Psychological Heterogeneity in User Preferences of Real-Time AI Mediation As Cognitive Scaffolds: A Latent Class Analysis

Anonymous CogSci submission

Abstract

Real-time AI mediation modifies, augments, or generates messages, requiring extensions of classical language processing theories to account for mediated delivery and timing in speech. While prior research has examined AI tools from a system-centered perspective, little is known about the heterogeneity of user profiles and discrete subgroup patterns in behavioral and perceptual responses to AI mediation. This study employed latent class analysis on qualitative interview data from 29 non-native English and Japanese speakers who interacted with a proactive AI assistant during collaborative conversations. Eight distinct two-class groupings emerged across thematic domains: Language and Background, Online Meeting Experience, AI Usage and Trust, Communication Strategies, Overall Evaluation, and feature-specific evaluations. Critically, class membership across domains showed minimal correspondence, indicating that user profiles reflect domain-specific adaptation patterns rather than a global typology. These findings revealed multidimensional user heterogeneity in adopting AI-mediation and provided a human-centered framework for understanding cognitive phenomena in human-AI interaction.

Keywords: AI-mediated communication; latent class analysis; cognitive scaffold; individual differences

Introduction

As artificial intelligence (AI) increasingly mediates communication, from smart suggestions to summaries, a critical question emerges: Why do some users benefit from AI assistance while others experience disruption? Empirical evidence reveals substantial heterogeneity in how people respond to AI-mediated communication (AI-MC): some over-rely on suggestions even when incorrect (Buçinca et al., 2021; Vasconcelos et al., 2023), others exhibit miscalibrated confidence or feel intruded by unsolicited assistance that impairs decision-making (Harari and Amir, 2025; Ma et al., 2024). Understanding the sources of this heterogeneity is a fundamental question about how cognitive processes, shaped by individual differences (ID) in processing capacity and accumulated experience, adapt when AI intervenes in real-time.

Cognitive Architecture of ID

IDs in cognitive processing are well-established across domains. Working memory capacity (WMC) predicts performance on tasks requiring controlled attention (Conway et al., 2003; Engle and Kane, 2004). High-WMC allows superior attentional control and maintains task goals despite distractions, while low-WMC relies more heavily on external cues (Barrett et al., 2004). ID encompass qualitatively distinct strategies in language processing (Kidd et al., 2018). Some develop direct, interlocutor-focused repair strategies to manage comprehension breakdowns, while others employ indirect, resource-based approaches (MacWhinney, 2005). These ID accumulate through experience, shaping linguistic competence and metacognitive beliefs about effective communication (Ortega, 2009). Critically, these differences cannot be fully captured by continuous measures or demographics alone. Latent class analysis (LCA) can reveal discrete subgroups with qualitatively distinct profiles that predict outcomes beyond what continuous variables explain (Collins and Lanza, 2010; Sinha et al., 2021). Yet human-centered approaches to understanding AI-MC remain limited.

When AI Helps Versus Hinders

AI assistance effects depend fundamentally on the match between what users need and what they can afford in real time (Sweller et al., 2019). When AI scaffolds align with the processing demands and WMC, they reduce extraneous load and enable deeper engagement. Just-in-time AI interventions after independent assessment achieved higher trust and preference (Kuang et al., 2024; Li et al., 2024). However, over-reliance occurs when users lack resources or motivation to critically evaluate, treating AI as a cognitive shortcut (Buçinca et al., 2021; Swaroop et al., 2025). Confidence miscalibration arises when users struggle to detect AI errors due to limited capacity (Ma et al., 2024). Unsolicited help undermines users' sense of competence and autonomy, reducing willingness to accept assistance and likelihood of future use (Harari and Amir, 2025), even without significantly degrading objective task performance (Mannem et al., 2023). These phenomena reflect interactions between user profiles and AI affordances, yet there lacks a theoretical framework to represent this heterogeneity in real-time AI-MC systematically, where cognitive demands fluctuate dynamically, and social coordination imposes additional constraints.

From Face-to-Face to AI-MC

In classical language production theories (e.g., Levelt, 1989), speakers produce disfluencies (i.e., fillers, pauses, repairs) as observable markers of internal processing challenges that signal to interlocutors and elicit collaborative repair (Arnold et al., 2007; Clark, 1996; Pickering and Garrod, 2013). Computer-mediated communication (CMC) altered these dynamics through reduced social cues and asynchrony, where AI mediation further introduces additional processing and evaluation, prompting theories of strategic adaptation (Hancock et al., 2020; Kiesler et al., 1984; Walther, 2011). Yet there lacks a systematic characterization of how users utilize these interventions in relation to their ID in cognitive profiles, and how the mismatch between their real-time states and intervention timing determines the interference of AI scaffolds.

Present Study

This study investigates user heterogeneity in AI-MC through qualitative interview data, asking: What discrete subgroup patterns characterize how users with diverse cognitive-linguistic profiles adapt to proactive AI scaffolds? We employ human-centered methods to identify qualitatively distinct user profiles across multiple domains: language backgrounds, mediation experiences, AI usage patterns, communication strategies, and feature evaluations. By identifying discrete subgroups with distinct processing patterns, strategic preferences, and evaluation criteria, we elucidate that effective AI-MC requires accommodating multidimensional user heterogeneity subject to WMC, affordance, and human-AI cognitive alignment in real time.

Methods

We conducted a controlled lab study with non-native speakers (NNS) of English and Japanese who represented diverse linguistic and cultural backgrounds. The study employed a Wizard-of-Oz method to integrate a proactive AI assistant into a Zoom-style interface, where participants engaged in a standardized conversation with a confederate.

Participants

Thirty-five NNS of English and Japanese (M=9, F=25, NA=1; M_{age}=20.79, SD=2.17; $M_{\text{years in English country}}$=7, SD=3.2) were recruited via the university system for course credit. Twenty-nine participants (M=7, F=19, NA=1; M_{age}=21.23, SD=2.25; $M_{\text{years in English country}}$=6.32, SD=2.95) completed the post-study interview.

Materials

AI-intervention Description We designed the sidebar in the Zoom interface using Figma with three common interventions: clarification, suggestion, and summary. To mimic Zoom interface, we included: 1) a caption box, where each green word is clickable with a pop-up box above that clarifies the term, and 2) a Zoom-style sidebar, where each feature pops up separately in a temporal order. The Zoom window was pinned to the placeholder, allowing participants to interact with the sidebar and caption at the same positions as in a real Zoom meeting. The tool was designed to be perceived as automated, with outputs generated proactively in real time.

Task The collaborative task conversation was to design menu items for a Japanese-American fusion restaurant in Collegetown for two reasons: 1) this topic was familiar to student groups to elicit genuinely engaging discussions, simulating task-based discussions; 2) the content was unfamiliar to NNS participants so they would intend to use AI scaffolds; and 3) we picked two cultures to avoid cases where some NNS may be familiar with American cuisine after years living in the US, and some NNS may be familiar with Japanese cuisine from personal interests. Validated through three pre-tests, this theme ensured a low level of topic and cultural familiarity for target participants from a third culture and allowed for naturalistic exchanges with open-ended questions.

Conversation and AI-interventions were standardized by pre-scripting the confederate part by ChatGPT to ensure consistency across participants. Confederates were trained to deliver the script naturally. Three English speakers revised the content, and pre-test feedback confirmed its naturalness and smoothness. Clarifications included 5 slang terms, 5 acronyms, 6 idioms, 6 complex words, and 6 cultural food items with images. Words within each category varied in relative complexity but roughly matched in lexical frequency in COCA (Davies, 2008). All words were positioned in the first half of the caption to ensure enough processing time. Clarification was restricted to twenty words. Six idea suggestions (3 American and 3 Japanese items) and four sentence suggestions (with bidirectional opinions to avoid bias) auto-popped up immediately after questions. Suggestion length was restricted to one full sentence. Four summaries in bullet points with high information density auto-popped up immediately after topic transitions. All features were only visible to participants and distributed evenly throughout the conversation.

Survey Participants completed a post-study survey on overall comprehensibility, perceived effectiveness, behavioral engagement, communication effort, and comfort with user interface (Duan et al., 2021). We also included questions about ID, including personality traits, language anxiety, alongside language background questions (He and Fussell, 2025).

Interview A semi-structured interview protocol was developed, asking participants to reflect on their experience using each feature. Additional questions were included to understand ID from their language acquisition backgrounds, prior use of AI tools, online meeting experiences, and current communication challenges and strategies.

Procedure

Two laptops joined over Zoom and connected in TeamViewer. The confederate monitored and progressed the conversation via TeamViewer. After consent, the participant joined a demo session to interact with features (3 mins). In the official session (15 mins), the confederate elicited a natural conversation with each user. They shared feedback via an interview (15-30 mins). All studies were conducted in English.

Data Analyses

Interviews were transcribed and manually reviewed by for errors. Using an inductive and interpretive approach (Glaser and Strauss, 2017), four coders identified the general themes, with high-level themes and their relationships further constructed. Codes were refined through an iterative process of creating and combining themes with mutual agreement.

Using the interview codes merged into eight higher-level thematic groups, we (1) identified distinct groups of users with similar patterns of perceptions and traits via latent class analysis (LCA), and (2) determined what factors predict which subgroups a person belongs to via model selection.

Each group was systematically evaluated with varying numbers of latent classes, beginning with two-class solutions and incrementally increasing to four-class solutions. Fit of individual models and subsequent model selection were conducted based on information criteria—Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), log-likelihood (ℓ), average posterior probability (AvgPP), and entropy (H). BIC was the primary selection criterion given its robust performance across diverse simulation conditions, while AIC provided complementary perspectives on model adequacy (Dziak et al., 2020). All LCA models were estimated using PROC LCA Version 1.2.5 (Lanza et al., 2007).

Within each group, each code's AvgPP was normalized for each latent class based on the model selected above. Codes were assigned to the class corresponding to their maximum probability. Risk differences (RD) measure the difference in normalized code rates between classes, where codes with $|\mathrm{RD}| \geq 0.20$ were interpreted as practically significant discriminators; using conventional effect heuristics, $|\mathrm{RD}| > 0.50$ was treated as a large effect, $|\mathrm{RD}|$=0.30–0.50 as medium, and $|\mathrm{RD}|$=0.10–0.30 as small (Higgins et al., 2019).

To examine whether latent class membership was associated with participant demographics and feature-specific perceptions, we conducted a series of statistical tests across eight LCA groupings derived from qualitative interview data. We assessed between-class differences in demographic variables using one-way ANOVAs for continuous variables and Fisher's exact tests for categorical variables. Additionally, we examined feature rating differences both between classes (using independent samples t-tests) and within classes (using repeated measures ANOVAs). To assess correspondence across LCA groupings, we conducted pairwise chi-square tests (or Fisher's exact tests when cell sizes < 5) across all 28 possible pairs of LCA groupings. Given the exploratory nature of this pattern characterization analysis and small sample size, we report p-values without Bonferroni correction to preserve the ability to detect real effects.

Results

Qualitative Characterization of User Subgroups

For each thematic group of interview codes, models with two through four latent classes were compared. The BIC values suggested that the two-class model was slightly superior, with AIC and AvgPP supporting the same model in most groups. All eight groups demonstrated excellent LCA model fit with AvgPP exceeding 0.90, indicating robust two-class solutions with minimal classification error of codes. We selected two-class models for all groups, given the greater parsimony, theoretical validity, and a good number of interview codes with above-medium RD in each class in terms of interpretability.

Language and Background Model comparisons favored a two-class solution (AIC=462.22, BIC=499.14, $\ell = -204.11$). Classification quality was perfect (H=0.00; $\mathrm{AvePP_{Class1}}$=$\mathrm{AvePP_{Class2}}$=1.00; $\mathrm{N_{Class1}} = 20$, $\mathrm{N_{Class2}} = 9$), though Class 1 predominated over Class 2. Class 1 exhibited a clustered group of Mandarin first language (L1) speakers (0.95 vs. 0.00; RD=0.95) alongside intermediate oral proficiency (0.25 vs. 0.00; RD=0.25) and a medium length of residence in the US (0.35 vs. 0.11; RD=0.24), emphasizing a *L1-dominant* profile. Class 2 showed higher rates for all other L1 speakers (0.00 vs. 1.00; RD= -1.00), and variations in living in the U.S. pre-college, reflecting mixed experience with a more *diverse-immersion* profile.

Online Meeting Experience Fit indices again supported two latent classes (AIC=210.81, BIC= 231.32, $\ell = -90.40$) with strong but skewed classification (H=0.01; $\mathrm{AvePP_{Class1}}$ =$\mathrm{AvePP_{Class2}}$ = 1.00; $\mathrm{N_{Class1}} = 22$, $\mathrm{N_{Class2}} = 7$). Class 1 exhibited more mentions of "internet problems" (0.73 vs. 0.00; RD=0.73), "audio lag" (0.50 vs. 0.00; RD=0.50), and "more miscommunications" (0.41 vs. 0.00; RD=0.41), reflecting a *technical-disruption* emphasis. Class 2 reported strong "online meeting motivations" (0.09 vs. 1.00; RD= -0.91) and higher satisfaction (0.00 vs. 0.43; RD= -0.43) but with "difficulty in technical collaborations" (0.05 vs. 0.43; RD= -0.38), defining an *engaged-user* profile.

AI Usage and Trust Model selection indicated two latent classes (AIC=472.05, BIC=511.70, $\ell = -207.03$). Classification was strong but slightly skewed (H=0.06; $\mathrm{AvePP_{Class1}}$=$\mathrm{AvePP_{Class2}}$=0.99; $\mathrm{N_{Class1}} = 18$, $\mathrm{N_{Class2}} = 11$). Class 1 had markedly higher probabilities for "AI use case for Academic Work" (0.67 vs. 0.00; RD=0.67), "diverse AI tools used" (0.61 vs. 0.18; RD=0.43), and "at least once per day" (0.44 vs. 0.18; RD=0.26), forming an *frequent-academic-use* profile. Class 2, by contrast, more often avoided AI tools in meetings ("never": 0.22 vs. 1.00; RD= -0.78), "low-less than once per week": 0.00 vs. 0.18; RD= -0.18), and showed lower trust for "mathematical tasks" (0.00 vs. 0.18; RD= -0.18), portraying a *infrequent-cautious* profile.

Communication Strategies A two-class model fit best (AIC=423.46, BIC=463.11, $\ell = -182.73$), with moderate classification (H=0.21; $\mathrm{AvePP_{Class1}}$=0.94; $\mathrm{AvePP_{Class2}}$=0.96; $\mathrm{N_{Class1}} = 17$, $\mathrm{N_{Class2}} = 12$). Class 1 predominantly favored in-moment repair, such as "ask person speaking for clarification on unfamiliar words/topics" (0.59 vs. 0.08; RD=0.50) or "attempt to answer unfamiliar question based on context" (0.35 vs. 0.08; RD=0.27), highlighting a *Direct-Solution* strategy. Class 2 leveraged external resources, such as "search unfamiliar questions online" (0.06 vs. 0.33; RD= -0.27), "compromise to avoid conflict" (0.00 vs. 0.25; RD= -0.25), or "ignore unfamiliar word/topic if not that important" (0.29 vs. 0.50; RD= -0.21), reflecting an *Indirect-Avoidance* approach.

Overall Evaluation For overall perceptions, a two-class solution was preferred (AIC=233.15, BIC=253.66, $\ell = -101.58$). Classification remained good (H=0.11; $\mathrm{AvePP_{Class1}}$=0.97; $\mathrm{AvePP_{Class2}}$=1.00; $\mathrm{N_{Class1}} = 20$, $\mathrm{N_{Class2}} = 9$). Class 1 exhibited higher probabilities for recognizing the helpfulness for "professional settings" (0.80 vs. 0.11; RD=0.69), "unfamiliar context/words" (0.70 vs. 0.33; RD=0.37), and "lecture" (0.35 vs. 0.00; RD=0.35), forming a *broad-adopter* profile. Class 2, however, included more instances of "AI unhelpful for interview" (0.10 vs.

0.00; RD=0.10), "lecture" (0.10 vs. 0.00; RD=0.10), and "delayed reaction in response" (0.10 vs. 0.00; RD=0.10), mapping to a *critical-evaluator* stance.

Clarification Evaluation Again, two classes emerged (AIC=608.19, BIC=669.71, $\ell = -259.09$), with adequate classification (H=0.06; $\text{AvePP}_{\text{Class1}}$=$\text{AvePP}_{\text{Class2}}$=0.99; $N_{\text{Class1}} = 7$, $N_{\text{Class2}} = 22$). Class 1 was characterized by "helpful in professional settings" (0.71 vs. 0.14; RD=0.58), "interviews" (0.57 vs. 0.00; RD=0.57), but not for "study groups" (0.57 vs. 0.18; RD=0.39), characterizing a *professional-performance* profile. Class 2 favored "lectures" (0.00 vs. 0.45; RD= −0.45), "helps conversation flow" (0.00 vs. 0.23; RD= −0.23), and "other alternatives may be used over feature" (0.00 vs. 0.18; RD= −0.18), defining an *academic-learner* profile. Aggregated results reinforce a distinction between professional and academic uses.

Suggestion Evaluation Suggestion also adhered to two classes (AIC=347.22, BIC=389.61, $\ell = -142.61$) with solid classification (H=0.10; $\text{AvePP}_{\text{Class1}}$=0.99; $\text{AvePP}_{\text{Class2}}$=0.97; $N_{\text{Class1}} = 13$, $N_{\text{Class2}} = 16$). Class 1 more frequently found suggestions "not helpful for study groups" (0.62 vs. 0.00; RD=0.62), "dependent on context" (0.62 vs. 0.00; RD=0.62), and "helpful for professional settings" (0.85 vs. 0.38; RD=0.47), indicating a *contextual-professional* profile. Class 2 more frequently included "concern about influence on response" (0.23 vs. 0.62; RD= −0.39), "helpful in general" (0.08 vs. 0.31; RD= −0.24), and "helpful for discussion meetings" (0.08 vs. 0.25; RD= –0.17), reflecting a *cautious-adopter* stance.

Summary Evaluation Summary confirmed a two-class fit (AIC=466.20, BIC=511.32, $\ell = -200.10$) with high quality (H=0.03; $\text{AvePP}_{\text{Class1}}$=0.93; $\text{AvePP}_{\text{Class2}}$=1.00; $N_{\text{Class1}} = 5$, $N_{\text{Class2}} = 24$). Class 1 exhibited high probabilities for use case in "large lectures" (1.00 vs. 0.25; RD=0.75), but not helpful for "study groups" (0.80 vs. 0.17; RD=0.63) and "interview due to time delay it causes" (0.60 vs. 0.04; RD=0.56), illustrating a *lecture-focused* adoption. Class 2 often showed "only helpful if lost in conversation" (0.00 vs. 0.17; RD= −0.17), "helpfulness varies" (0.00 vs. 0.12; RD= −0.12), and "not enough time to read" (0.00 vs. 0.12; RD= −0.12), reflecting a *situational-adopter* profile.

User Profiles: Between-Group Associations

Only Communication Strategies was significantly associated with length of immersion ($F(1,25)$ =4.91, p=.036). $\text{Communication}_{\text{Class1}}$ (*Direct-Solution*) had significantly longer immersion (M=7.4 yrs, SD=2.9) compared to $\text{Communication}_{\text{Class2}}$ (*Indirect-Avoidance*; M=5.0 yrs, SD=2.7). This finding suggests that communication strategy preferences identified through interviews relate systematically to language immersion experience. More extensive immersion predicts greater reliance on direct interlocutor-focused repair mechanisms (asking for clarification, using context) rather than external compensatory strategies (online searches, avoidance, compromise). All other demographic comparisons across the eight groupings were non-significant. The sample was approximately 71% female across all groupings, with no significant gender differences in class membership distributions.

Within-Group Feature Differentiation

A robust pattern emerged when examining whether participants differentiated among the four feature ratings: 10 of 16 class-level analyses showed significant within-class feature differentiation, with participants consistently rating Overall experience higher than Summary. The consistent Overall > Summary hierarchy (significant in 7 of 10 post-hoc comparisons) suggests that, regardless of how users conceptualized and discussed AI scaffolds qualitatively, they converged on a shared quantitative perception that holistic impressions of the tool outweigh the evaluations of the Summary feature.

One exception emerged when examining the relationship between feature ratings and class membership: high Summary raters ($\geq$ median) were significantly more likely to belong to $\text{Language/Background}_{\text{Class2}}$ (*Diverse-Immersion*) compared to low Summary raters ($\chi^2(1)$=6.21, p=.013; 57.1% vs. 7.1% in $\text{Language/Background}_{\text{Class2}}$). This suggests that linguistic diversity and varied immersion backgrounds may have cultivated cognitive flexibility, enabling more effective summary processing.

Discussion

This study employed latent class analysis to identify discrete subgroup patterns in user profiles, behavioral adaptations, and perceptual responses to AI-MC scaffolds. Rather than evaluating the proactive AI system itself, this analysis revealed fundamental insights about user heterogeneity that extend cognitive theories from F2F contexts to AI-MC, illuminate shifts in communication patterns under AI mediation, and inform adaptive design principles for diverse user populations.

Domain-Specific Adaptation

Participants did not consistently cluster across all eight groupings, indicating that adaptation to AI-MC is multidimensional rather than unitary. This pattern challenges traditional technology adoption models that assume a single underlying dimension (e.g., "digital literacy") manifests uniformly across contexts. Our findings align with multidimensional models, recognizing different user profiles along distinct dimensions: linguistic dominance, prior AI experience, cognitive offloading strategies, and repair approaches.

This multidimensionality has critical theoretical implications for extending F2F models to AI-MC. Classical theories of language production posit that disfluencies serve as observable markers of internal processing states that have multiple options of responding to delays and errors, with different disfluency types reflecting what strategies are available given the processing demands (Arnold et al., 2007; Clark, 1996; Levelt, 1989). In AI-MC, users bring distinct profiles that determine which processing demands they experience and how they allocate cognitive resources. For instance, users favoring *Direct-Solution* may perceive AI clarifications as redundant for interlocutor-repair mechanisms they already deploy effectively, whereas *Indirect-Avoidance* users may treat the same clarifications as novel cognitive scaffolds that substitute for strategies they lack. The psycholinguistic framework must therefore account for ID in baseline repair strategies when predicting how AI mediation redistributes cognitive load.

Similarly, theories of cognitive offloading and external scaffolding have largely assumed homogeneous user populations. The emergence of distinct AI Usage and Trust classes (Frequent-Academic-Use vs. Infrequent-Cautious) demonstrates that users differ fundamentally in their propensity to offload cognitive tasks to AI systems. This heterogeneity aligns with recent work showing that technology adoption reflects not merely perceived usefulness and ease of use but also IDs in cognitive abilities, personality traits, and domain-specific trust. The present findings show that offloading propensity itself constitutes a discrete user profile dimension, independent of other user characteristics.

Language Immersion as Accumulated Capital

The only significant demographic predictor was years of English-speaking country experience, supporting accumulated adaptation rather than static traits of AI-MC patterns. This pattern aligns with second language acquisition theories, emphasizing that immersion provides linguistic input and sociocultural competence in navigating communication breakdowns (MacWhinney, 2005; Ortega, 2009). Through repeated exposure to F2F interactions where misunderstandings arise, NNS with longer immersion develop a more sophisticated linguistic repertoire, metalinguistic awareness, and confidence in deploying social repair mechanisms. When these users encounter AI scaffolds, they interpret them through existing repair strategies rather than as novel affordances.

From a cognitive architecture perspective, this finding suggests communication strategies in AI-MC reflect learned heuristics stored in long-term memory and activated automatically based on situational cues, rather than being computed anew in WM for each interaction. Participants with extensive immersion have accumulated richer schemas for managing communication challenges, including knowledge about when to seek clarification directly versus when to rely on contextual inference (A. D. Baddeley and Hitch, 1974; Sweller et al., 2019). These schemas guide their engagement with AI scaffolds, determining whether proactive clarifications are perceived as helpful support or unnecessary interruption.

The independence of demographic predictors reinforces the domain-specificity of AI-MC adaptation. Language immersion predicted communication strategy profiles specifically, but not AI usage patterns, online meeting experiences, or feature evaluations. This dissociation suggests different dimensions of AI-MC adaptation draw on distinct knowledge bases: linguistic capital for communication repair, technological familiarity for AI trust and usage, and task-specific experience for feature evaluation.

Universal Cognitive Load Patterns

Feature ratings revealed a robust universal pattern: Overall Tool Experience was consistently rated higher than the Summary feature, regardless of users' qualitative profiles. This dissociation likely reflects cognitive load dynamics (Sweller, 1988; Sweller et al., 2019). Summaries present high information density in compressed format during ongoing conversation, imposing substantial dual-task demands: comprehending summary content while maintaining conversational flow. Participants may have experienced summary integration as extraneous cognitive load that exceeded WMC (A. Baddeley, 2003), particularly in real-time contexts where germane load (deep processing) was already allocated to formulating responses. Expertise reversal effect may have further amplified Summary-related cognitive load (Chen et al., 2017). Users with stronger comprehension skills or more developed discourse schemas may have found summaries redundant, introducing extraneous information that interfered with their own mental representations of the conversation structure. The exception that high Summary raters were disproportionately in Language/Background$_{\text{Class2}}$ (*Diverse-Immersion*) suggests that linguistic diversity may have cultivated cognitive flexibility, enabling more effective dual-task management.

This universal cognitive load pattern demonstrates that certain aspects of language processing constrain AI-MC experiences regardless of ID in strategies, usage patterns, or background. Limitations of WMC impose upper bounds on how much proactive assistance can be presented simultaneously in real-time interaction (Kirschner et al., 2011).

Timing, Proactivity, and Receptivity

While real-time scaffolds aim to provide timely support during moments of processing difficulty, their effectiveness can be undermined by the misalignment among intervention timing, users' dynamic cognitive states, and their psychological receptivity to unsolicited intervention. Our findings reveal timing-dependent receptivity: *Direct-Solution* strategists perceived clarifications as redundant when they resolve difficulties spontaneously, while *Indirect-Avoidance* users needed them for genuine comprehension difficulty. Yet real-time conversation lacks discrete problem boundaries where optimal timing is identifiable. The universal cognitive load pattern (Overall $>$ Summary) provides indirect evidence: Introducing compressed informational summaries at precisely these high-demand moments may represent a double mismatch: poor timing and unsolicited assistance.

These findings contribute to growing evidence that effective proactive AI-MC requires: (1) timing models identifying receptive moments, (2) initiative strategies preserving autonomy and competence, and (3) personalization accounting for IDs in WMC and strategic preferences.

Extending Theory from F2F to AI-MC

Our findings revealed that adaptation to AI-MC involves multiple independent dimensions: **Linguistic capital transfer** determines how language backgrounds and immersion histories shape interpretation of AI-generated content. **Mediation competence** reflects prior experiences with technical disruptions versus engaged platform usage. **Cognitive offloading propensity** captures willingness to delegate tasks based on accumulated usage and trust. **Adaptive strategies** encompass repair mechanisms developed through F2F and CMC experience that influence receptivity to AI scaffolds. **Context-dependent evaluation** reflects domain-specific mental models determining when AI support aligns with task demands.

Critically, the independence of these dimensions challenges models predicting AI-MC adaptation from single factors like digital literacy, suggesting modular cognitive processes engaged during AI-MC. Users appear to recruit different cognitive and metacognitive strategies depending on whether they are evaluating features (analytical processes), deploying repairs (social processes), or assessing trust (accumulated experience). Future theoretical work should integrate situated cognition frameworks that emphasize how environmental affordances (e.g., AI scaffolds) interact with users' existing competencies and mental models (Risko and Gilbert, 2016). Models should account for how ID in baseline processing strategies, accumulated linguistic capital, and cognitive flexibility moderate the impact of AI scaffolds.

Understanding Individual Differences in Cognition

Traditional ID research focused on continuous variables like WMC, verbal fluency, or vocabulary as predictors of comprehension and production (Conway et al., 2003; Kidd et al., 2018). LCA can reveal discrete subgroups with qualitatively distinct patterns not captured by continuous measures alone.

Communication Strategies (*Direct-Solution* and *Indirect-Avoidance*) represent qualitatively different orientations toward managing communication challenges: one emphasizes social engagement and in-the-moment negotiation, while the other prioritizes external resources and conflict avoidance. These orientations likely reflect not only linguistic competence but also personality traits, cultural norms around directness, and meta-cognitive beliefs about effective communication. Similarly, the AI Usage and Trust dimension revealed discrete profiles that likely reflect accumulated experience shaping mental models of AI capabilities and appropriate delegation. These profiles may be more predictive of how users engage with novel AI systems than continuous measures of digital literacy, as they capture specific usage patterns and trust calibration developed over time.

Methodological Contribution

This study demonstrates LCA's value for bridging qualitative depth and quantitative structure. While our inductive codes rely on interpretive consensus rather than formal reliability metrics (a limitation for purely quantitative validation), LCA's probabilistic framework is uniquely suited to model the inherent "fuzziness" of such qualitative categories. By treating code patterns as probabilistic signals rather than deterministic classifications, LCA reveals structural regularities in domain-independent ways that emerge despite potential variations in codings. This dissociation between qualitative heterogeneity and quantitative universality addresses longstanding tensions between idiographic (individual) and nomothetic (universal) approaches (Kidd et al., 2018), offering a scalable pathway to systematize rich, unstructured user data in cognitive science.

Limitations and Future Directions

First, NNS student sample limits the generalizability to other populations. The skewed gender distribution, though not predictive of class membership, may affect representativeness. The WoZ methodology sacrificed ecological validity for consistent experimental control. Second, the study examined a single AI system with specific features. Future research should investigate cognitive mechanisms underlying these discrete profiles through eye-tracking and neuroimaging, experimentally manipulate AI intervention timing to test receptivity windows in communication contexts, examine how proactive versus reactive AI assistance differentially affects user profiles, and link profiles to behavioral outcomes (i.e., communication effectiveness, learning gains, relationship quality) to establish practical significance.

Conclusion

User adaptation to AI-MC is multidimensional, domain-specific, and cognitively constrained. Eight distinct user profile dimensions captured heterogeneity with minimal cross-domain correspondence. Language immersion emerged as a significant predictor of communication strategy profiles, reflecting accumulated adaptation through experience. Quantitative feature ratings revealed a universal Overall > Summary hierarchy, reflecting cognitive load dynamics in real-time conversation. These findings extend language processing theories, showing that modular AI-MC adaptation requires recognition of discrete subgroup patterns. The domain-specificity challenges unitary technology adoption models and informs understanding of how speech management, WM allocation, and strategic adaptation operate when intelligent agents proactively intervene. Critically, the universal cognitive load pattern, effectiveness of timing-dependent interventions, and psychological costs of unsolicited assistance suggest that AI-MC systems should move toward user-adaptive models that account for individual processing profiles, dynamic cognitive states, psychological receptivity to proactive support, and contextually appropriate timing. By shifting focus from tool evaluation to user profile heterogeneity, this work illuminates fundamental questions about human cognitive flexibility, the limits of WMC under augmented interaction, and the theoretical extensions necessary as communication increasingly involves real-time AI collaboration.

Acknowledgments

We used ChatGPT to generate the study material (pre-scripted conversation and features), reformat the equations in Results into LaTeX codes, and revise the grammatical errors.

References

Arnold, J. E., Kam, C. L. H., & Tanenhaus, M. K. (2007). If you say thee uh you are describing something hard: The on-line attribution of disfluency during reference comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(5), 914.

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature reviews neuroscience*, *4*(10), 829–839.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.). Academic Press. https://doi.org/https://doi.org/10.1016/S0079-7421(08)60452-1

Barrett, L. F., Tugade, M. M., & Engle, R. W. (2004). Individual differences in working memory capacity and dual-process theories of the mind. *Psychological bulletin*, *130*(4), 553.

Buçinca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To trust or to think: Cognitive forcing functions can reduce overreliance on ai in ai-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), 1–21. https://doi.org/10.1145/3449287

Chen, O., Kalyuga, S., & Sweller, J. (2017). The expertise reversal effect is a variant of the more general element interactivity effect. *Educational Psychology Review*, *29*(2), 393–405.

Clark, H. H. (1996). *Using language*. Cambridge university press.

Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences*. John Wiley & Sons.

Conway, A. R., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in cognitive sciences*, *7*(12), 547–552.

Davies, M. (2008). The corpus of contemporary american english. www.english-corpora.org/coca/

Duan, W., Yamashita, N., Shirai, Y., & Fussell, S. R. (2021). Bridging fluency disparity between native and nonnative speakers in multilingual multiparty collaboration using a clarification agent. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW2), 1–31. https://doi.org/10.1145/3479579

Dziak, J. J., Coffman, D. L., Lanza, S. T., Li, R., & Jermiin, L. S. (2020). Sensitivity and specificity of information criteria. *Briefings in bioinformatics*, *21*(2), 553–565.

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of. *The psychology of learning and motivation: Advances in research and theory*, *44*, 145.

Glaser, B. G., & Strauss, A. L. (2017). *The discovery of grounded theory: Strategies for qualitative research*. Routledge.

Hancock, J. T., Naaman, M., & Levy, K. (2020). Ai-mediated communication: Definition, research agenda, and ethical considerations. *Journal of Computer-Mediated Communication*, *25*(1), 89–100. https://doi.org/10.1093/jcmc/zmz022

Harari, D., & Amir, O. (2025). Proactive ai adoption can be threatening: When help backfires. *arXiv preprint arXiv:2509.09309*.

He, W., & Fussell, S. R. (2025). Difference in the Cognitive Mechanism of Predictive Processing in Computer-Mediated Communication: A Comparison Study of L2 Speakers. *Proceedings of the 43rd Annual Conference of the Cognitive Science Society*.

Higgins, J. P., Li, T., & Deeks, J. J. (2019). Choosing effect measures and computing estimates of effect. *Cochrane handbook for systematic reviews of interventions*, 143–176.

Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Individual differences in language acquisition and processing. *Trends in cognitive sciences*, *22*(2), 154–169.

Kiesler, S., Siegel, J., & McGuire, T. W. (1984). Social psychological aspects of computer-mediated communication. *American psychologist*, *39*(10), 1123.

Kirschner, F., Paas, F., & Kirschner, P. A. (2011). Task complexity as a driver for collaborative learning efficiency: The collective working-memory effect. *Applied Cognitive Psychology*, *25*(4), 615–624.

Kuang, E., Li, M., Fan, M., & Shinohara, K. (2024). Enhancing ux evaluation through collaboration with conversational ai assistants: Effects of proactive dialogue and timing. *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3613904.3642168

Lanza, S. T., Collins, L. M., Lemmon, D. R., & Schafer, J. L. (2007). Proc lca: A sas procedure for latent class analysis. *Structural equation modeling: a multidisciplinary journal*, *14*(4), 671–694.

Levelt, W. J. (1989). *Speaking: From intention to articulation*. MIT press.

Li, C., Zhang, Z., Saugstad, M., Safranchik, E., Kulkarni, C., Huang, X., Patel, S., Iyer, V., Althoff, T., & Froehlich, J. E. (2024). Labelaid: Just-in-time ai interventions for improving human labeling quality and domain knowledge in crowdsourcing systems. *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 1–21.

Ma, S., Wang, X., Lei, Y., Shi, C., Yin, M., & Ma, X. (2024). "are you really sure?" understanding the effects of human self-confidence calibration in ai-assisted decision making. *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 1–20.

MacWhinney, B. (2005). A unified model of language acquisition (J. F. Kroll & A. De Groot, Eds.). *Handbook of bilingualism: Psycholinguistic approaches*, *4967*, 50–70.

Mannem, S., Macke, W., Stone, P., & Mirsky, R. (2023). Exploring the cost of interruptions in human-robot teaming. *2023 IEEE-RAS 22nd International Conference on Humanoid Robots (Humanoids)*, 1–8.

Ortega, L. (2009). *Understanding second language acquisition*. Hodder Education.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329–347. https://doi.org/10.1017/S0140525X12001495

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences*, *20*(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002

Sinha, P., Calfee, C. S., & Delucchi, K. L. (2021). Practitioner's guide to latent class analysis: Methodological considerations and common pitfalls. *Critical care medicine*, *49*(1), e63–e79.

Swaroop, S., Buçinca, Z., Gajos, K. Z., & Doshi-Velez, F. (2025). Personalising ai assistance based on overreliance rate in ai-assisted decision making. *Proceedings of the 30th International Conference on Intelligent User Interfaces*, 1107–1122.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive science*, *12*(2), 257–285.

Sweller, J., Van Merriënboer, J. J., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational psychology review*, *31*(2), 261–292.

Vasconcelos, H., Jörke, M., Grunde-McLaughlin, M., Gerstenberg, T., Bernstein, M. S., & Krishna, R. (2023). Explanations can reduce overreliance on ai systems during decision-making. *Proceedings of the ACM on Human-Computer Interaction*, *7*(CSCW1), 1–38.

Walther, J. B. (2011). Theories of computer-mediated communication and interpersonal relations. *The handbook of interpersonal communication*, *4*, 443–479.

Table 1: LCA Feature Grouping Models

Group	Class	AIC	BIC	LogLik(ℓ)	Entropy(H)	AvgPP	Participants
AI Usage & Trust	**2**	**472.05**	**511.70**	**-207.03**	**0.0564**	**0.9903**	**18 / 11**
	3	474.31	534.47	-193.15	0.0786	0.9746	14 / 9 / 6
	4	482.86	563.53	-182.43	0.0219	0.9914	4 / 8 / 14 / 3
Clarification Eval	**2**	**608.19**	**669.71**	**-259.09**	**0.0631**	**0.9869**	**7 / 22**
	3	623.60	716.57	-243.80	0.0000	1.0000	5 / 3 / 21
	4	641.56	765.98	-229.78	0.0106	0.9973	6 / 13 / 1 / 9
Comm. Strategies	**2**	**423.46**	**463.11**	**-182.73**	**0.2148**	**0.9481**	**17 / 12**
	3	437.37	497.53	-174.68	0.0943	0.9559	18 / 8 / 3
	4	447.40	528.07	-164.70	0.0751	0.9564	3 / 2 / 11 / 13
Lang. & Background	**2**	**462.22**	**499.14**	**-204.11**	**0.0000**	**1.0000**	**20 / 9**
	3	461.11	517.17	-189.56	0.0163	0.9953	9 / 11 / 9
	4	463.14	538.34	-176.57	0.0166	0.9868	6 / 6 / 9 / 8
Online Meeting Exp	**2**	**210.81**	**231.32**	**-90.40**	**0.0058**	**0.9994**	**22 / 7**
	3	215.54	246.99	-84.77	0.0000	1.0000	21 / 5 / 3
	4	225.89	268.28	-81.95	0.0001	1.0000	13 / 8 / 3 / 5
Overall Evaluation	**2**	**233.15**	**253.66**	**-101.58**	**0.1072**	**0.9797**	**20 / 9**
	3	239.87	271.31	-96.93	0.0399	0.9869	7 / 17 / 5
	4	243.96	286.34	-90.98	0.0749	0.9580	4 / 10 / 3 / 12
Suggestion Eval	**2**	**347.22**	**389.61**	**-142.61**	**0.0984**	**0.9758**	**13 / 16**
	3	368.56	432.82	-137.28	0.1112	0.9564	13 / 12 / 4
	4	384.91	471.05	-129.46	0.0001	1.0000	1 / 1 / 16 / 11
Summary Evaluation	**2**	**466.20**	**511.32**	**-200.10**	**0.0318**	**0.9884**	**5 / 24**
	3	477.95	546.31	-188.97	0.0298	0.9875	12 / 10 / 7
	4	496.92	588.52	-181.46	0.0044	0.9986	2 / 3 / 3 / 21

Table 2: Participant Class Memberships

ID	Lang/Back	Online Exp	AI Use	Comm Strat	Overall	Clarify	Suggest	Summary
C1-8	1	2	1	1	1	2	1	2
C2-3	1	1	2	2	1	1	2	1
C3-3	1	2	2	1	1	2	1	1
C5-9	1	1	1	1	2	2	2	2
C6-3	1	1	2	1	1	2	1	2
Ki7-4	2	2	1	2	1	1	1	2
V9-3	1	1	1	2	1	2	1	1
A10-3	2	1	1	2	1	2	1	2
C11-9	1	1	2	1	1	2	2	2
C14-3	1	1	1	1	1	2	2	2
K15-5	2	2	1	1	1	2	2	2
C16-11	1	1	1	2	2	2	2	2
K17-4	2	1	1	2	2	1	2	2
C18-5	1	1	1	2	1	2	2	2
C19-3	1	1	1	2	2	2	2	2
C21-5	1	1	2	2	2	2	2	2
U22-10	2	1	2	1	1	2	1	2
C23-6	1	1	1	1	1	2	1	2
C24-11	1	1	1	1	1	2	2	2
K25-1	2	1	1	1	2	1	1	2
G27-11	2	1	2	1	2	2	2	2
C28-6	1	1	2	1	1	2	1	2
C29-11	1	2	1	1	2	1	1	2
C30-5	1	2	1	2	1	2	2	2
P31-4	2	2	2	1	1	1	1	1
C32-9	1	1	2	2	2	2	2	2
C33-7	1	1	1	1	1	1	1	1
H34-9	2	1	1	1	1	2	2	2
C35-5	1	1	2	2	1	2	2	2

Note: Participants were denoted by their native language, participant number, and years spent in an English-speaking country (e.g., C3-3=3rd Chinese-speaking participant, 3 years. First letter of Participant ID denotes native language: C=Chinese, G=German, H=Hindi, K=Korean, Ki=Kinyarwanda, P=Persian, U=Urdu, V=Vietnamese, A=Unknown.

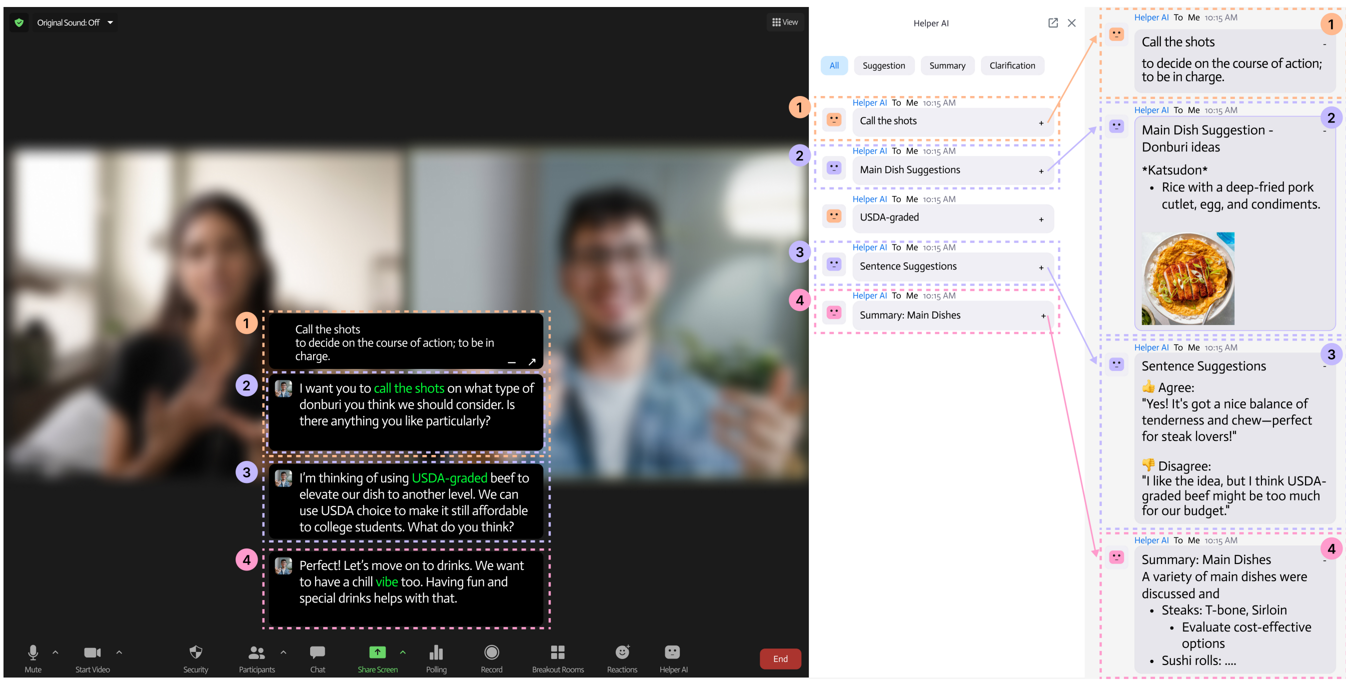

Figure 1: Zoom interface with XPLAIN; 1: clarification box popped up above caption once clicked with minimal eye movement track, saved as minimized in the side bar for future reference; 2: idea suggestion; 3: full-sentence suggestion; 4: summary. 2 3 4 all popped up as expanded windows. Only one feature expands at a time to avoid visual overload. (N.B. the caption stacks were for illustration purposes: only one caption box stayed at the bottom position as the conversation progressed.)

Figure 2: Top Codes from Each Latent Class Per Group. Each plot represents a group, where the five most prominent codes from each latent class are included with absolute values of risk differences.