

Original Paper

Large Language Model–Based Simplification of Digital Therapeutics Explanations for Insomnia and Nicotine Dependence: Two Randomized Online Experiments

JunYoung Seo¹, MS; Moses Yook¹, MS; Dai Jin Kim², MD; JunHee Lee², MD; Jae Hyun Yoo², MD; GiHwan Byeon², MD; In Young Choi¹, PhD

¹Department of Medical Informatics, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea

²Department of Psychiatry, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea

Corresponding Author:

In Young Choi, PhD

Department of Medical Informatics, College of Medicine

The Catholic University of Korea

222, Banpo-daero, Seocho-gu

Seoul

Republic of Korea

Phone: 82 2-3147-8423

Email: iychoi@catholic.ac.kr

Abstract

Background: Digital therapeutics (DTx) are evidence-based software interventions with the potential to treat health conditions. However, uptake remains limited by low public awareness and overly complex patient education materials that exceed recommended readability levels. Large language models (LLMs) may simplify such content; however, their effects on users' understanding have not been empirically demonstrated.

Objective: This study aimed to examine whether LLM-based simplification of DTx explanatory materials enhances perceived understanding and subjective evaluations of readability, clarity, and comprehensibility compared with manufacturer-provided documents.

Methods: We developed a simplification tool using the GPT-4o application programming interface (API), configured for deterministic outputs and guided by structured readability instructions. Original DTx explanatory materials about insomnia and nicotine dependence were obtained from manufacturers and transformed into simplified versions. Two randomized, between-subject online experiments were conducted (n=1000, with 500 participants in each experiment). Participants were stratified by age and sex and screened for relevance (Insomnia Severity Index ≥ 8 for the insomnia experiment and smoking ≥ 5 cigarettes per day for the nicotine dependence experiment). Within each experiment, participants were randomly assigned to review either the original or the LLM-simplified explanatory material. Perceived understanding was assessed before and after exposure. Postexposure evaluations of ease, clarity, and comprehensibility were also collected.

Results: Repeated measures ANOVA revealed significant group \times time interaction effects on perceived understanding in both experiments: insomnia ($F_{1,498}=24.8$; $P<.001$) and nicotine dependence ($F_{1,498}=14.1$; $P<.001$), with greater improvements observed in the LLM-simplified groups. Mann-Whitney U tests further showed that LLM-simplified explanations were rated as easier, clearer, and more comprehensible than the original versions in both experiments (all $P<.05$), with small to moderate effect sizes ($r=0.11-0.24$).

Conclusions: Compared with manufacturer-provided original materials, LLM-simplified DTx explanations led to greater improvements in perceived understanding and subjective evaluations of readability among lay audiences, even after a single exposure. This finding highlights the potential scalability of LLM-based simplification as a strategy to improve the perceived accessibility of health information for lay audiences. Integrating such tools into patient education may enhance how lay audiences perceive and engage with DTx, although further research using objective comprehension measures is needed to confirm these benefits.

Trial Registration: Clinical Research Information Service KCT0011459; <https://cris.nih.go.kr/cris/search/detailSearch.do?seq=32221>

Keywords: digital therapeutics; large language models; text simplification; patient education materials; randomized online experiment

Introduction

Background

Digital therapeutics (DTx) represent a significant advancement in health care and are defined as high-quality software programs delivering evidence-based therapeutic interventions for a wide range of physical, mental, and behavioral conditions [1-3]. Unlike traditional pharmaceuticals and devices, DTx are often cost-effective and designed to encourage sustained user engagement, making them a promising avenue for modern health care [4,5]. DTx have been applied to various conditions, including chronic pain, cardiometabolic diseases (eg, diabetes and hypertension), heart failure, and mental health disorders. They demonstrate potential to complement pharmacotherapy; improve adherence; enable remote monitoring; and expand therapeutic options through emerging technologies, such as virtual reality and wearable sensors [6-10]. In this study, we focused on 2 DTx use cases, insomnia and nicotine dependence, as accessible and representative examples to evaluate the effects of large language model (LLM)-based simplification.

Prior Work

Despite promising applications, the adoption of DTx remains limited [11,12]. Public awareness is low, and many individuals outside the medical community are unfamiliar with DTx or do not fully understand their purpose [13]. For new health technologies to achieve broader societal diffusion, individuals must develop a clear understanding of their value [14-16].

A significant barrier is the inherent complexity of medical information. Research indicates that the average health literacy level of US adults corresponds to that of 13- to 14-year-olds, and medical education materials are recommended to be written at the level of 10- to 11-year-olds [17]. However, over 95% of existing patient education materials require a reading level of at least high school, making them inaccessible to a significant portion of the general population [18]. Producing simplified yet accurate resources remains a resource-intensive and time-consuming process that requires substantial expertise [19-21].

LLMs offer a promising solution. Trained on vast amounts of text data, LLMs can perform natural language tasks, such as summarization, translation, and simplification [22]. Their ability to adapt content to different comprehension levels has been shown to improve the readability of health information, often quantified through automated metrics, such as Flesch-Kincaid grade levels [23-25]. However, these evaluations have rarely examined whether improved readability translates to greater comprehension or more positive user perceptions among lay audiences. This gap highlights the need for empirical studies that directly assess how individuals

experience and interpret simplified LLM-generated health information.

Study Objectives

This study aimed to address these gaps by empirically testing whether LLM-based simplification of DTx explanatory materials for insomnia and nicotine dependence enhances perceived understanding and subjective evaluations. Two outcome categories were assessed. The first was the cognitive outcome, which examined whether exposure to LLM-simplified explanations yields greater perceived understanding than exposure to the original explanations. The second was the subjective outcome, which examined whether LLM-based simplification improves evaluations of explanatory materials regarding ease, clarity, and overall comprehensibility.

Specifically, we aimed to determine whether LLM-simplified DTx explanations, compared with the original manufacturer-provided materials, lead to (1) greater improvements in perceived understanding and (2) higher subjective ratings of ease, clarity, and comprehensibility, across 2 independent randomized online experiments.

Methods

Study Design

This study used a randomized, between-subjects online experimental design to evaluate the effectiveness of LLM-simplified DTx explanations. Two independent experiments were conducted, each focusing on a different therapeutic area: insomnia or nicotine dependence.

In both experiments, participants were randomly assigned to one of two groups:

1. Control group—received the original manufacturer-provided DTx explanatory materials
2. Experimental group—received LLM-simplified versions of the same materials

This design allowed direct comparison of the effects of the original versus simplified explanations on participants' perceived understanding and subjective evaluations.

Development of the LLM-based Simplification Tool

To address the challenge of making DTx information more comprehensible to the public while maintaining clinical accuracy, we developed a lightweight text simplification tool. The tool used the GPT-4o (OpenAI) application programming interface (API), a state-of-the-art LLM, for its advanced natural language processing capabilities. The primary goal was to generate simplified text that would be understandable to lay audiences while ensuring fidelity to the original medical content.

In alignment with recommendations for health education materials that suggest a readability level suitable for 10- to 11-year-olds [17], the tool was specifically designed to simplify complex DTx explanations to this level, ensuring accessibility across diverse populations regardless of health literacy.

The simplification process was guided by structured instructions covering content preservation, logical coherence, terminology adaptation, clarity enhancement, tone and style, emphasis, length management, and readability. The full prompt text provided to the LLM is available in [Multimedia Appendix 1](#).

For model configuration, the temperature parameter was set to 0 to minimize randomness and maximize reproducibility, and the top p parameter was fixed at 1 to ensure comprehensive consideration of relevant tokens. This deterministic setup was critical for maintaining precision and consistency in the medical information. These design choices enabled the creation of a tool that converts complex DTx documents into simplified, patient-friendly explanations without compromising medical accuracy.

DTx Explanatory Materials

To evaluate the generalizability of the proposed text simplification approach, we selected 2 DTx that addressed distinct health conditions: insomnia and nicotine dependence. This design enabled the assessment of the applicability of the LLM-based simplification tool across different therapeutic domains, rather than within a single disease context.

For each DTx, original patient education and explanatory documents were obtained directly from the respective manufacturers. These materials, characterized by high complexity and technical terminology, served as baseline texts for the simplification process. Using the LLM-based tool described above, simplified versions of each document were generated, resulting in two parallel sets of materials for each DTx product: (1) the original manufacturer-provided version and (2) the LLM-simplified version.

The LLM-simplified versions incorporated several modifications, including vocabulary simplification, tone adjustment, and formatting improvements, such as clearer paragraph structuring and emphasis of key points. In contrast, the original manufacturer-provided materials were presented in their original plain-text format without any modification.

To ensure medical accuracy and logical coherence, all LLM-generated materials were independently reviewed by a panel of 3 board-certified psychiatrists who regularly treat conditions such as insomnia and nicotine dependence. This expert validation confirmed that the simplified content faithfully preserved the original medical meaning, avoided factual errors or misinterpretations, and presented information in a logically coherent manner. This review process was essential to maintaining the clinical integrity and trustworthiness of the materials used in this study. The full texts of the original and LLM-simplified explanatory materials are provided in [Multimedia Appendix 2](#).

Participants and Recruitment

In total, 1000 participants were recruited through an online panel service (Macromill Embrain Co, Ltd) [26], with 500 participants allocated to each experiment (n=500, 50% to insomnia DTx and n=500, 50% to nicotine dependence DTx). Stratified assignment was applied during recruitment to ensure comparable sex and age distributions between the control and experimental groups.

The general inclusion criteria were (1) age between 19 and 64 years, (2) access to a smartphone or computer with internet connectivity, and (3) provision of informed consent before participation. Condition-specific screening thresholds were established in consultation with 3 board-certified psychiatrists to enhance contextual relevance while recruiting participants from the general population rather than from patient groups. For the insomnia DTx experiment, only participants with an Insomnia Severity Index score of ≥ 8 were included, corresponding to at least mild insomnia symptoms [27]. The Insomnia Severity Index is a 7-item self-report scale assessing insomnia symptom severity, with total scores ranging from 0 to 28. For the nicotine dependence DTx experiment, only participants who reported smoking at least 5 cigarettes per day were eligible. These criteria were selected to ensure that participants, although not recruited as patients, shared health-related characteristics that made the DTx materials meaningful and relevant to their personal contexts.

Procedure

This study used an online, randomized, between-subjects experimental design. Participants were recruited through an online panel service and screened according to the general and condition-specific inclusion criteria described above. Only individuals who met eligibility requirements were invited to participate.

After confirming eligibility, participants provided digital informed consent. In each experiment, participants were recruited from an online panel service [26] that collected demographic information (age and sex) during the screening process. To ensure balance across conditions, stratified randomization was applied, assigning participants to either the control group (original materials) or the experimental group (LLM-simplified materials) within each demographic stratum. This approach minimized allocation bias and maintained between-group comparability.

Following the group assignment, participants completed a structured online questionnaire in the following sequence:

- Pre-exposure assessment of perceived understanding of the respective DTx
- Exposure to the assigned explanatory materials (original vs LLM-simplified) with a minimum enforced viewing time of 30 seconds
- Postexposure assessment of perceived understanding of the respective DTx
- Evaluation of explanations, where participants rated the difficulty, clarity, and overall comprehensibility of the materials

All responses were anonymized to ensure the strict protection of participant privacy. After completing the survey, participants received compensation according to the agreement established by the online panel service.

Measures

Perceived Understanding

Participants' perceived understanding of the 2 DTx was assessed both before and after exposure to the explanatory materials. This construct was measured using 5 items adapted from Flynn and Goldsmith subjective knowledge scale [28], modified to capture how much participants believed they understood each DTx. Responses were recorded on a 5-point Likert scale ranging from 0 (not at all) to 4 (very much). The exact wording of the 5 adapted items is provided in the [Multimedia Appendix 3](#). The scale demonstrated good internal consistency, with Cronbach $\alpha=0.80$ for the insomnia DTx experiment and $\alpha=0.82$ for the nicotine dependence DTx experiment.

Evaluation of Explanations

Participants' evaluation of the explanatory materials was measured after exposure using single-item ratings on a 0 to 4 scale, where 0 indicated "not at all" and 4 indicated "very much." Three aspects were assessed: (1) ease (how easy or difficult the explanation was to read and follow), (2) clarity (how clear the explanation was perceived overall), and (3) comprehensibility (the extent to which participants felt they could understand the explanation).

All questionnaires were administered in Korean. No objective knowledge tests (eg, recall or factual questions) were included; thus, all outcome measures relied on participants' subjective self-reports.

Ethical Considerations

This study was approved by the Institutional Review Board of the Catholic Medical Center (approval MC25QNSI0039) and was conducted in accordance with the principles of the Declaration of Helsinki [29]. All participants provided informed consent before participation. All participant data

were anonymized and kept strictly confidential; no personally identifiable information was collected or retained. Participants received compensation for their participation in accordance with the terms established by the online panel service [26]. This study is reported in accordance with the Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth (CONSORT-EHEALTH) guidelines; the completed checklist is provided in [Checklist 1](#).

Statistical Analysis

All statistical analyses were conducted using Jamovi (version 2.6.26). A significance threshold of $\alpha=.05$ was applied to all tests. To examine the effect of explanation type on perceived DTx understanding, repeated measures ANOVA was conducted. Each model tested the group (LLM-simplified vs control) \times time (before vs after exposure) interaction to determine whether changes in perceived understanding over time significantly differed between groups. The assumptions of normality, homoscedasticity, and sphericity were satisfied.

Participants' postexposure ratings of ease, clarity, and comprehensibility of the DTx explanations were compared between groups using the Mann-Whitney *U* test. This nonparametric approach was selected because of the ordinal nature of the single-item Likert-scale ratings. These analyses evaluated whether LLM-based simplification had a significant influence on participants' subjective evaluations of the materials.

Results

Sample Characteristics

A total of 1000 participants completed the study, with 500 (50%) allocated to the insomnia DTx experiment and 500 (50%) to the nicotine dependence DTx experiment. Stratified assignment ensured comparable sex and age distributions between the control and experimental groups. Demographic information used for stratification was limited to sex and age. Participant demographic characteristics are presented in [Table 1](#).

Table 1. Participant demographic characteristics by group (insomnia and nicotine dependence experiments).

Characteristic	Insomnia control (n=250)	Insomnia test (n=250)	Nicotine control (n=250)	Nicotine test (n=250)
Sex, n (%)				
Male	125 (50)	125 (50)	125 (50)	125 (50)
Female	125 (50)	125 (50)	125 (50)	125 (50)
Age (years), n (%)				
20	63 (25.2)	63 (25.2)	63 (25.2)	63 (25.2)
30	63 (25.2)	63 (25.2)	63 (25.2)	63 (25.2)
40	62 (24.8)	62 (24.8)	62 (24.8)	62 (24.8)
50-65	62 (24.8)	62 (24.8)	62 (24.8)	62 (24.8)

Effect on Perceived Understanding

No significant main effect of group was observed in either the insomnia experiment ($F_{1,498}=3.32$; $P=.07$; partial $\eta^2=0.007$)

or the nicotine dependence experiment ($F_{1,498}=2.88$; $P=.09$; partial $\eta^2=0.006$; [Table 2](#)).

Table 2. Repeated measures ANOVA results for perceived understanding across insomnia and nicotine dependence digital therapeutics (DTx) experiments.

Experiment and effect	F test (df)	P value	Partial η^2
Insomnia DTx			
Group	3.32 (1, 498)	.07	0.007
Time	38.5 (1, 498)	<.001	0.072
Group×time	24.8 (1, 498)	<.001	0.048
Nicotine DTx			
Group	2.88 (1, 498)	.09	0.006
Time	21.5 (1, 498)	<.001	0.041
Group×time	14.1 (1, 498)	<.001	0.028

A significant main effect of time was observed in both the insomnia experiment ($F_{1,498}=38.5$; $P<.001$; partial $\eta^2=0.072$) and the nicotine dependence experiment ($F_{1,498}=21.5$, $P<.001$; partial $\eta^2=0.041$), with postexposure assessment scores higher than pre-exposure assessment scores across groups.

A significant group×time interaction was observed in both experiments. In the insomnia experiment, the interaction

effect was $F_{1,498}=24.8$; $P<.001$; partial $\eta^2=0.048$; and in the nicotine dependence experiment, it was $F_{1,498}=14.1$; $P<.001$; partial $\eta^2=0.028$ (Table 2).

Descriptive statistics for each group and time point are presented in Table 3. As shown in Figure 1, the LLM-simplified group demonstrated greater increases in perceived understanding than the control group in both experiments.

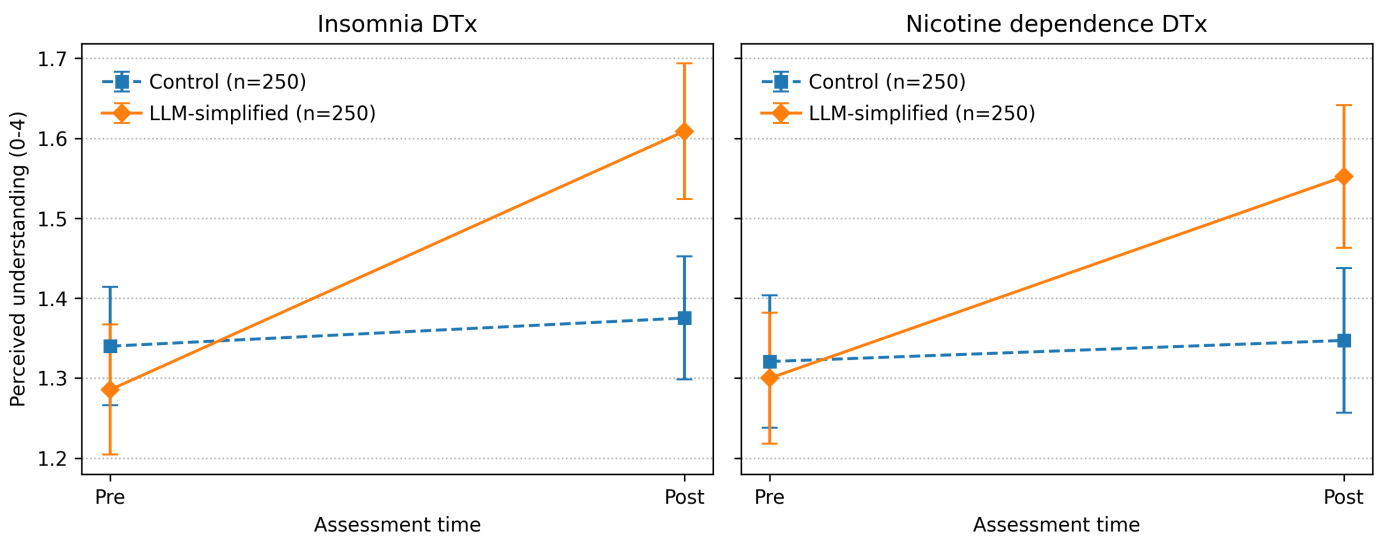
Table 3. Perceived understanding scores by group and time point (n=250).

Experiment and group	Pretest, mean (SD; 95% CI)	Posttest, mean (SD; 95% CI)
Insomnia DTx ^a		
LLM ^b -simplified	1.29 (0.65; 1.20-1.37)	1.61 (0.68; 1.52-1.69)
Control	1.34 (0.59; 1.27-1.41)	1.38 (0.62; 1.30-1.45)
Nicotine DTx		
LLM-simplified	1.30 (0.66; 1.22-1.38)	1.55 (0.72; 1.46-1.64)
Control	1.32 (0.66; 1.24-1.40)	1.35 (0.73; 1.26-1.44)

^aDTx: digital therapeutics.

^bLLM: large language model.

Figure 1. Changes in perceived digital therapeutics (DTx) understanding from pre-exposure assessment to postexposure assessment in both experiments. Points indicate group means, and error bars represent 95% CIs. LLM: large language model.



Evaluation of Explanations

In the insomnia experiment, the LLM-simplified explanations were rated significantly higher for ease ($U=27,281$; $P=.008$; $r=0.127$), clarity ($U=27,803$; $P=.02$; $r=0.110$), and comprehensibility ($U=25,125$; $P<.001$; $r=0.196$).

Similarly, in the nicotine dependence experiment, significant group differences were observed for ease ($U=23,727$; $P<.001$; $r=0.240$), clarity ($U=27,697$; $P=.02$; $r=0.114$), and comprehensibility ($U=24,738$; $P<.001$; $r=0.208$). The distribution of ratings across response categories is illustrated in Figure 2.

Figure 2. Group distribution of postexposure ratings (0-4) for ease, clarity, and comprehensibility in both experiments. Bars represent 100% stacked proportions. The neutral category (2) is split and plotted on both sides of the center, such that lower ratings (0-1) extend to the left of 0 and higher ratings (3-4) extend to the right. This split-neutral presentation allows the direction of responses (negative vs positive) to be visually compared while preserving the full distribution of ratings. Group labels indicate sample sizes (control vs large language model [LLM]–simplified). Between-group differences were tested using 2-sided Mann-Whitney U tests with Holm correction across measures; rank-biserial effect sizes (r) are reported in Table 4. DTx: digital therapeutics.

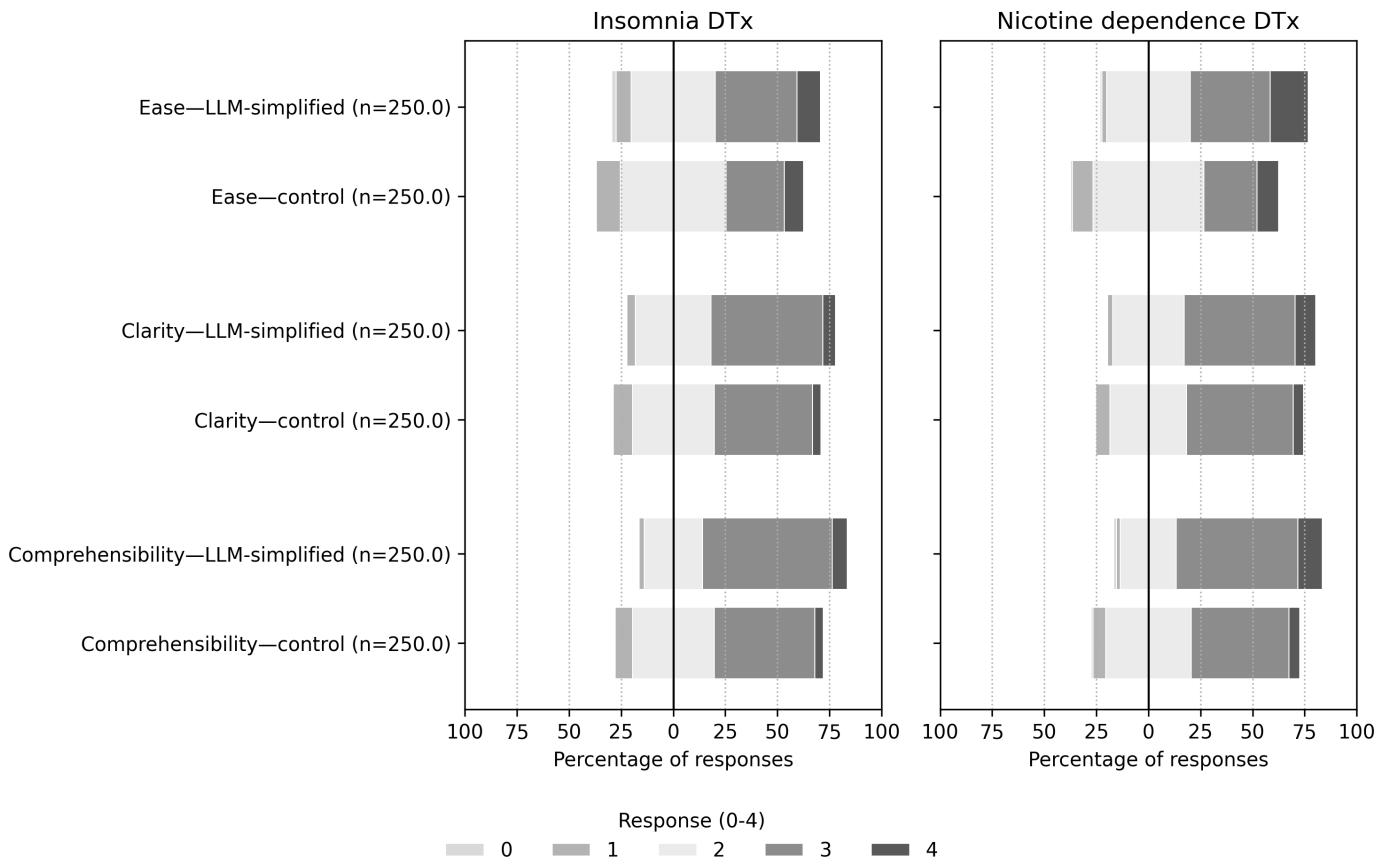


Table 4. Mann-Whitney U test results comparing subjective evaluation of explanations between groups.

Experiment and measure	U	P value	Adjusted P value ^a	r
Insomnia DTx^b				
Ease	27,281	.008	.02	0.127
Clarity	27,803	.02	.03	0.11
Comprehensibility	25,125	<.001	.006	0.196
Nicotine DTx				
Ease	23,727	<.001	.006	0.24
Clarity	27,697	.02	.03	0.114
Comprehensibility	24,738	<.001	.006	0.208

^aHolm correction applied for multiple comparisons.

^bDTx: digital therapeutics.

Discussion

This study evaluated whether LLM-based simplification of DTx explanatory materials improves perceived understanding and subjective evaluations among lay audiences through 2 independent randomized online experiments targeting insomnia (500/1000, 50%) and nicotine dependence (500/1000, 50%).

Principal Results

Our findings indicated that LLM-based simplification of DTx explanatory materials improves perceived comprehensibility and subjective evaluations of accessibility. In 2 independent experiments on insomnia and nicotine dependence, participants who received LLM-simplified explanations reported significantly greater gains in perceived understanding than those who received the original manufacturer-provided explanations. In addition, the simplified versions were consistently rated as easier, clearer, and more comprehensible, indicating improved subjective evaluations. These findings suggest that LLM-based simplification may improve perceived understanding and subjective evaluations across multiple therapeutic contexts. However, further research using objective comprehension measures and more diverse populations is needed to confirm the robustness and generalizability of these findings.

Although these findings demonstrate improvements in perceived understanding, these improvements do not necessarily indicate accurate or complete comprehension of medical information. Feeling confident about understanding health information may differ from actually understanding the details and appropriate use of a medical intervention. Therefore, future research should incorporate objective knowledge measures to determine whether LLM-based simplification improves actual comprehension in addition to perceived understanding.

The observed differences in subjective evaluations were statistically significant, but the effect sizes were generally small to moderate ($r=0.11-0.24$). These magnitudes suggest modest but consistent improvements in participants' evaluations following a single brief exposure. Given the minimal intervention required, such improvements may still be meaningful in real-world settings where scalable approaches to improving the perceived accessibility of health information are needed.

These findings emphasize the potential for safe and practical applications of LLMs in health care communication. Although concerns about hallucinations highlight the risks of using LLMs to generate new clinical content or make medical decisions [30], their role in transforming clinically validated information into more accessible formats carries considerably lower risk [31]. Prior research has demonstrated that readability can be improved while maintaining factual accuracy [23-25]. This study extends that evidence by incorporating independent reviews from 3 board-certified psychiatrists who confirmed the medical accuracy and logical coherence of the simplified texts. Collectively, these results

demonstrate that LLM-based simplification represents a scalable and trustworthy strategy for improving the perceived accessibility and comprehensibility of DTx information among lay audiences.

Comparison With Prior Work

The findings of this study align with a substantial body of research that highlights the challenges of health literacy and the frequent mismatch between the readability of medical information and patients' comprehension skills [17]. Previous studies have consistently shown that most patient education materials are written at a level above the recommended reading level, limiting accessibility for the general population [18]. Recent advances in natural language processing and LLMs suggest that automated text simplification can serve as a promising solution, with evidence demonstrating improvements in readability across a range of health-related documents [22]. However, much of this prior work relied on computational readability metrics (such as Flesch-Kincaid scores) rather than empirical evaluations with end users [23-25].

This study extends this line of research by providing experimental evidence that LLM-simplified explanations not only achieve higher readability but also enhance perceived understanding among lay audiences. Notably, the results demonstrate that accuracy can be preserved, consistent with prior findings that LLM-based simplification improves accessibility without compromising factual correctness [23-25]. Unlike earlier studies, this study combines expert clinical validation by board-certified psychiatrists with large-scale user experiments across 2 therapeutic contexts, thereby offering novel empirical evidence of both accuracy and effectiveness. These findings reinforce the growing recognition that LLM-based simplification may help improve the perceived accessibility of DTx information and support users' perceived understanding of complex health technologies.

Limitations

This study has some limitations. First, participants were recruited from an online panel and allocated through stratified randomization based on sex and age, which may have reduced natural variability and limited the generalizability of the results [32]. Because participation required internet access and familiarity with digital environments, the study sample may have had higher digital literacy than typical patient populations. Therefore, the findings may not fully generalize to individuals with limited digital literacy. In addition, because demographic information was limited to sex and age, other potentially relevant characteristics (such as education, income, health literacy, and prior DTx exposure) were not collected, further limiting the interpretability of the results. In particular, differences in education level and baseline health literacy may influence how individuals respond to simplified health information and should be considered when interpreting the generalizability of the findings. Second, comprehension was assessed only through self-reported perceived understanding rather than objective knowledge tests [33], which limits conclusions about actual

comprehension. Third, ease, clarity, and comprehensibility were evaluated using single-item measures, restricting measurement reliability. Fourth, the short-term exposure with a minimum enforced viewing time of 30 seconds leaves the durability and behavioral relevance of the effects uncertain. Fifth, although participants were screened for relevance (eg, minimum insomnia symptoms or daily smoking frequency), they were not clinically diagnosed; thus, caution is warranted when extrapolating these findings to treatment populations. Sixth, the intervention relied on a single LLM (GPT-4o) and specific prompt configurations, underscoring the need for replication using alternative models and implementation conditions to establish robustness. Finally, the intervention involved multiple simultaneous modifications, including vocabulary simplification, tone adjustment, and formatting improvements, such as paragraph structuring and emphasis of key points. Notably, the original manufacturer-provided materials were presented in plain text without visual enhancements, whereas the LLM-simplified versions incorporated formatting elements, such as bolding, structured paragraph breaks, and emphasis of key terms, as part of the simplification process. Therefore, the comparison was not purely linguistic; differences in visual presentation may have independently contributed to the observed improvements in perceived comprehensibility and subjective evaluations. This constitutes a bundled intervention, and the present design does not allow for disentangling the effects of linguistic simplification from those of visual formatting. Future studies should isolate these components—for example, through factorial designs that independently vary linguistic complexity and visual formatting—to better understand their individual and combined effects.

Conclusions

This study demonstrates that LLM-based simplification of DTx explanatory materials significantly improves perceived understanding and subjective evaluations of readability, clarity, and comprehensibility among lay audiences. By focusing on the transformation of clinically validated content, LLMs can provide a safe and scalable approach to improving the perceived accessibility of digital health information.

Beyond DTx, these findings have broader implications: many types of medical documents, including patient education materials, informed consent forms, and clinical instructions, can be made to feel more accessible through LLM-based simplification. Importantly, by helping patients feel more confident in engaging with health information, LLM-based tools may have the potential to reduce passive compliance and instead promote active engagement, adherence, and shared decision-making.

Future studies should extend these results in several ways. Research involving clinically diagnosed patient populations is necessary to confirm the applicability of this approach in real-world treatment settings. Additionally, objective knowledge assessments and longitudinal designs are needed to determine whether improvements in perceived understanding translate into sustained learning and behavioral change. Finally, evaluations across diverse health conditions, populations, and languages are essential to establish the robustness and generalizability of LLM-based simplification strategies.

Acknowledgments

The authors thank Aimnext Co, Ltd and Innerwave Co, Ltd for their cooperation in providing the original explanatory materials for the digital therapeutics evaluated in this study.

Funding

This work was supported by the Bio-Industrial Technology Development Program (RS-2022-KT217960; Development of an online platform for DTx full-cycle clinical trial support using metaverse technology) funded by the Ministry of Trade, Industry and Resources (MOTIR, Korea). This funding covered both the research process and the article processing charge. The funder had no role in the design of the study, data collection, analysis, interpretation, or manuscript preparation.

Data Availability

The datasets generated and/or analyzed during this study are not publicly available due to participant privacy considerations but are available from the corresponding author on reasonable request, subject to renewed institutional review board approval.

Authors' Contributions

Conceptualization: JYS (lead), IYC (supporting)

Data curation: JYS (lead)

Formal analysis: JYS (lead), MY (supporting)

Funding acquisition: DJK (lead), IYC (supporting)

Investigation: JYS (lead)

Methodology: JYS (lead), MY (supporting)

Supervision: IYC (lead), MY (supporting)

Validation: JHY (equal), JHL (equal), GHB (equal; clinical consultation and review of the large language model-simplified materials)

Writing—original draft: JYS (lead)

Writing—review and editing: JYS (lead), IYC (supporting), MY (supporting), DJK (supporting), JHY (supporting), JHL (supporting), GHB (supporting)
All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Full prompt text used for the large language model.

[[DOCX File \(Microsoft Word File\), 18 KB-Multimedia Appendix 1](#)]

Multimedia Appendix 2

Original and large language model–simplified explanatory materials.

[[DOCX File \(Microsoft Word File\), 19 KB-Multimedia Appendix 2](#)]

Multimedia Appendix 3

Survey questionnaire.

[[DOCX File \(Microsoft Word File\), 19 KB-Multimedia Appendix 3](#)]

Checklist 1

CONSORT-EHEALTH (V 1.6.1) checklist for reporting randomized trials of eHealth interventions.

[[PDF File \(Adobe File\), 1296 KB-Checklist 1](#)]

References

1. Hong JS, Wasden C, Han DH. Introduction of digital therapeutics. *Comput Methods Programs Biomed.* Sep 2021;209:106319. [doi: [10.1016/j.cmpb.2021.106319](https://doi.org/10.1016/j.cmpb.2021.106319)] [Medline: [34364181](https://pubmed.ncbi.nlm.nih.gov/34364181/)]
2. Palanica A, Docktor MJ, Lieberman M, Fossat Y. The need for artificial intelligence in digital therapeutics. *Digit Biomark.* 2020;4(1):21-25. [doi: [10.1159/000506861](https://doi.org/10.1159/000506861)] [Medline: [32399513](https://pubmed.ncbi.nlm.nih.gov/32399513/)]
3. Digital Therapeutics Alliance. URL: <https://dtxalliance.org/> [Accessed 2025-04-28]
4. Li Y, Gong Y, Zheng B, et al. Effects on adherence to a mobile app-based self-management digital therapeutics among patients with coronary heart disease: pilot randomized controlled trial. *JMIR Mhealth Uhealth.* Feb 15, 2022;10(2):e32251. [doi: [10.2196/32251](https://doi.org/10.2196/32251)] [Medline: [34906924](https://pubmed.ncbi.nlm.nih.gov/34906924/)]
5. Sapanel Y, Tadeo X, Brenna CT, et al. Economic evaluation associated with clinical-grade mobile app-based digital therapeutic interventions: systematic review. *J Med Internet Res.* Aug 1, 2023;25:e47094. [doi: [10.2196/47094](https://doi.org/10.2196/47094)] [Medline: [37526973](https://pubmed.ncbi.nlm.nih.gov/37526973/)]
6. Biskupiak Z, Ha VV, Rohaj A, Bulaj G. Digital therapeutics for improving effectiveness of pharmaceutical drugs and biological products: preclinical and clinical studies supporting development of drug + digital combination therapies for chronic diseases. *J Clin Med.* Jan 11, 2024;13(2):403. [doi: [10.3390/jcm13020403](https://doi.org/10.3390/jcm13020403)] [Medline: [38256537](https://pubmed.ncbi.nlm.nih.gov/38256537/)]
7. Rohaj A, Bulaj G. Digital therapeutics (DTx) expand multimodal treatment options for chronic low back pain: the nexus of precision medicine, patient education, and public health. *Healthcare (Basel).* May 18, 2023;11(10):1469. [doi: [10.3390/healthcare11101469](https://doi.org/10.3390/healthcare11101469)] [Medline: [37239755](https://pubmed.ncbi.nlm.nih.gov/37239755/)]
8. Singh B, Hazra P, Roy S, et al. Exploring the need and benefits of digital therapeutics (DTx) for the management of heart failure in India. *Cureus.* Nov 2023;15(11):e49628. [doi: [10.7759/cureus.49628](https://doi.org/10.7759/cureus.49628)] [Medline: [38161874](https://pubmed.ncbi.nlm.nih.gov/38161874/)]
9. Nordyke RJ, Appelbaum K, Berman MA. Estimating the impact of novel digital therapeutics in type 2 diabetes and hypertension: health economic analysis. *J Med Internet Res.* Oct 9, 2019;21(10):e15814. [doi: [10.2196/15814](https://doi.org/10.2196/15814)] [Medline: [31599740](https://pubmed.ncbi.nlm.nih.gov/31599740/)]
10. Ferreira-Brito F, Ferreira J, Vieira M, Guerreiro J, Guerreiro T. Digital therapeutics with virtual reality and sensors. In: *UbiComp/ISWC '23 Adjunct: Adjunct Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing & the 2023 ACM International Symposium on Wearable Computing.* Association for Computing Machinery; 2023.
11. Armeni P, Polat I, De Rossi LM, Diaferia L, Meregalli S, Gatti A. Exploring the potential of digital therapeutics: an assessment of progress and promise. *Digit Health.* 2024;10:20552076241277441. [doi: [10.1177/20552076241277441](https://doi.org/10.1177/20552076241277441)] [Medline: [39291152](https://pubmed.ncbi.nlm.nih.gov/39291152/)]
12. Survey on perception of digital therapeutic devices. Health Insurance Review & Assessment Service. 2024. URL: <https://repository.hira.or.kr/handle/2019.oak/3208> [Accessed 2026-06-02]
13. Sim B, Ju JH, Kim B, Lee JY. Stakeholders' perceptions regarding digital therapeutics reimbursement in South Korea: qualitative study. *JMIR Mhealth Uhealth.* Oct 30, 2023;11:e47407. [doi: [10.2196/47407](https://doi.org/10.2196/47407)] [Medline: [37933414](https://pubmed.ncbi.nlm.nih.gov/37933414/)]

14. Feder G, Slade R. The acquisition of information and the adoption of new technology. *Am J Agric Econ*. Aug 1984;66(3):312-320. [doi: [10.2307/1240798](https://doi.org/10.2307/1240798)]
15. Straub ET. Understanding technology adoption: theory and future directions for informal learning. *Rev Educ Res*. Jun 2009;79(2):625-649. [doi: [10.3102/0034654308325896](https://doi.org/10.3102/0034654308325896)]
16. Tverskoi D, Babu S, Gavrillets S. The spread of technological innovations: effects of psychology, culture and policy interventions. *R Soc Open Sci*. Jun 2022;9(6):211833. [doi: [10.1098/rsos.211833](https://doi.org/10.1098/rsos.211833)] [Medline: [35754991](https://pubmed.ncbi.nlm.nih.gov/35754991/)]
17. Weiss BD. Health literacy: a manual for clinicians. American Medical Association Foundation and American Medical Association. 2003. URL: <http://lib.ncfh.org/pdfs/6617.pdf> [Accessed 2025-04-28]
18. Morony S, Flynn M, McCaffery KJ, Jansen J, Webster AC. Readability of written materials for CKD patients: a systematic review. *Am J Kidney Dis*. Jun 2015;65(6):842-850. [doi: [10.1053/j.ajkd.2014.11.025](https://doi.org/10.1053/j.ajkd.2014.11.025)] [Medline: [25661679](https://pubmed.ncbi.nlm.nih.gov/25661679/)]
19. Dickinson DK, McCabe A. Bringing it all together: the multiple origins, skills, and environmental supports of early literacy. *Learn Disabil Res Pract*. Nov 2001;16(4):186-202. [doi: [10.1111/0938-8982.00019](https://doi.org/10.1111/0938-8982.00019)]
20. Gal I. Towards probability literacy for all citizens: building blocks and instructional dilemmas. In: Jones GA, editor. *Exploring Probability in School*. Springer; 2005.
21. Sheppard ED, Hyde Z, Florence MN, McGwin G, Kirchner JS, Ponce BA. Improving the readability of online foot and ankle patient education materials. *Foot Ankle Int*. Dec 2014;35(12):1282-1286. [doi: [10.1177/1071100714550650](https://doi.org/10.1177/1071100714550650)] [Medline: [25239196](https://pubmed.ncbi.nlm.nih.gov/25239196/)]
22. Abd-Alrazaq A, AlSaad R, Alhuwail D, et al. Large language models in medical education: opportunities, challenges, and future directions. *JMIR Med Educ*. Jun 1, 2023;9:e48291. [doi: [10.2196/48291](https://doi.org/10.2196/48291)] [Medline: [37261894](https://pubmed.ncbi.nlm.nih.gov/37261894/)]
23. Kirchner GJ, Kim RY, Weddle JB, Bible JE. Can artificial intelligence improve the readability of patient education materials? *Clin Orthop Relat Res*. Nov 1, 2023;481(11):2260-2267. [doi: [10.1097/CORR.0000000000002668](https://doi.org/10.1097/CORR.0000000000002668)] [Medline: [37116006](https://pubmed.ncbi.nlm.nih.gov/37116006/)]
24. Rouhi AD, Ghanem YK, Yolchieva L, et al. Can artificial intelligence improve the readability of patient education materials on aortic stenosis? A pilot study. *Cardiol Ther*. Mar 2024;13(1):137-147. [doi: [10.1007/s40119-023-00347-0](https://doi.org/10.1007/s40119-023-00347-0)] [Medline: [38194058](https://pubmed.ncbi.nlm.nih.gov/38194058/)]
25. Patel EA, Fleischer L, Filip P, et al. The use of artificial intelligence to improve readability of otolaryngology patient education materials. *Otolaryngol Head Neck Surg*. Aug 2024;171(2):603-608. [doi: [10.1002/ohn.816](https://doi.org/10.1002/ohn.816)] [Medline: [38751109](https://pubmed.ncbi.nlm.nih.gov/38751109/)]
26. Embrain. URL: <https://www.embrain.com> [Accessed 2026-05-02]
27. Bastien CH, Vallières A, Morin CM. Validation of the Insomnia Severity Index as an outcome measure for insomnia research. *Sleep Med*. Jul 2001;2(4):297-307. [doi: [10.1016/s1389-9457\(00\)00065-4](https://doi.org/10.1016/s1389-9457(00)00065-4)] [Medline: [11438246](https://pubmed.ncbi.nlm.nih.gov/11438246/)]
28. Flynn LR, Goldsmith RE. A short, reliable measure of subjective knowledge. *J Bus Res*. Sep 1999;46(1):57-66. [doi: [10.1016/S0148-2963\(98\)00057-5](https://doi.org/10.1016/S0148-2963(98)00057-5)]
29. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA*. Nov 27, 2013;310(20):2191-2194. [doi: [10.1001/jama.2013.281053](https://doi.org/10.1001/jama.2013.281053)] [Medline: [24141714](https://pubmed.ncbi.nlm.nih.gov/24141714/)]
30. Bélisle-Pipon JC. Why we need to be careful with LLMs in medicine. *Front Med (Lausanne)*. 2024;11:1495582. [doi: [10.3389/fmed.2024.1495582](https://doi.org/10.3389/fmed.2024.1495582)] [Medline: [39697212](https://pubmed.ncbi.nlm.nih.gov/39697212/)]
31. Bojić L, Zagovora O, Zelenkauskaitė A, et al. Comparing large language models and human annotators in latent content analysis of sentiment, political leaning, emotional intensity and sarcasm. *Sci Rep*. Apr 3, 2025;15(1):11477. [doi: [10.1038/s41598-025-96508-3](https://doi.org/10.1038/s41598-025-96508-3)] [Medline: [40181141](https://pubmed.ncbi.nlm.nih.gov/40181141/)]
32. Bethlehem J. Selection bias in web surveys. *Int Stat Rev*. Aug 2010;78(2):161-188. [doi: [10.1111/j.1751-5823.2010.00112.x](https://doi.org/10.1111/j.1751-5823.2010.00112.x)]
33. Waters EA, Biddle C, Kaphingst KA, et al. Examining the interrelations among objective and subjective health literacy and numeracy and their associations with health knowledge. *J Gen Intern Med*. Nov 2018;33(11):1945-1953. [doi: [10.1007/s11606-018-4624-2](https://doi.org/10.1007/s11606-018-4624-2)] [Medline: [30120636](https://pubmed.ncbi.nlm.nih.gov/30120636/)]

Abbreviations

API: application programming interface

CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth

DTx: digital therapeutics

LLM: large language model

Edited by Yuri Quintana; peer-reviewed by Armaun Rouhi, Osama Loubani; submitted 23.Dec.2025; final revised version received 01.May.2026; accepted 02.May.2026; published 10.Jun.2026

Please cite as:

Seo J, Yook M, Kim DJ, Lee J, Yoo JH, Byeon GH, Choi IY

Large Language Model–Based Simplification of Digital Therapeutics Explanations for Insomnia and Nicotine Dependence: Two Randomized Online Experiments

JMIR Hum Factors 2026;13:e89451

URL: <https://humanfactors.jmir.org/2026/1/e89451>

doi: [10.2196/89451](https://doi.org/10.2196/89451)

© JunYoung Seo, Moses Yook, Dai Jin Kim, JunHee Lee, Jae Hyun Yoo, GiHwan Byeon, In Young Choi. Originally published in JMIR Human Factors (<https://humanfactors.jmir.org>), 10.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on <https://humanfactors.jmir.org>, as well as this copyright and license information must be included.