*healthcare*

MDPI

*Article*

# Factors Associated with Intended Professional Use of Artificial Intelligence Among Speech-Language Therapy Students and Practicing Clinicians: A Cross-Sectional Survey in Türkiye

**Şaziye Seçkin Yılmaz [1,*], Beyza Özen [2] and Alperen Özen [3]**

[1] Department of Speech and Language Therapy, Faculty of Health Sciences, Fenerbahçe University, Atatürk Mah., Ataşehir Bulvarı, Metropol İstanbul, Ataşehir, İstanbul 34758, Türkiye
[2] Gelişim Akademi Special Education and Rehabilitation Center, Ankara 06824, Türkiye; byza.ozenn@outlook.com
[3] Yeni Hayatlar Special Education and Rehabilitation Center, Ankara 06370, Türkiye; oezenalperen@gmail.com
* Correspondence: saziye.seckin@fbu.edu.tr

## Abstract

**Background/Objectives:** Little evidence shows which appraisals are associated with intended professional use of generative artificial intelligence (AI) in speech-language therapy. This study examined factors independently associated with intended AI use among final-year students and practicing clinicians in Türkiye. **Methods:** In a cross-sectional online survey (September–October 2026), participants completed a Turkish adaptation of a US-developed questionnaire. After excluding 66 of 443 respondents failing an attention check, the sample comprised 140 students and 237 clinicians, most of whom were early-career. Likelihood of AI use was regressed on six conceptually specified variables, with bootstrap and post hoc latent-variable and sensitivity analyses. **Results:** Perceived usefulness ($\beta = 0.427$) and experience with ChatGPT/AI ($\beta = 0.239$) were independently associated with likelihood of AI use ($R^2 = 0.359$); self-rated familiarity with AI, frequency of use, and comfort were not. A small positive association with replacement concern was unstable across specifications. After correction, students and clinicians differed on 1 of 11 dimensions. Most dimensions showed weak single-factor structure. **Conclusions:** Intended AI use was associated with perceived usefulness and direct experience, whereas self-rated familiarity, measured with low internal consistency, was not; the data do not support conclusions about the relative importance of experience and knowledge, and the cross-sectional design precludes causal inference.

**Keywords:** artificial intelligence; ChatGPT; speech-language therapy; health professions education; technology acceptance; intended use; continuing professional development; cross-sectional survey

Check for updates

Academic Editors: Daniele Giansanti and Joaquim Carreras

Received: 20 October 2026
Revised: 9 November 2026
Accepted: 17 November 2026
Published: 22 November 2026

**Copyright:** © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

## 1. Introduction

Generative artificial intelligence (AI) tools such as ChatGPT are now accessible to health professions students and clinicians, and their uptake has become the subject of a dedicated adoption literature [1,2]. In health professions education, medical students have reported positive perceptions of ChatGPT for treatment guidance and medical education, and both students and physicians rated it positively for producing patient education materials [3], while graduating medical students reported positive expectations alongside limited formal preparation [4]. A recent survey of endocrinologists and postgraduate trainees likewise

described heterogeneous AI exposure, frequent self-directed access to AI tools, and concerns about accuracy, overreliance, and privacy, with comparisons between trainees and specialists reported descriptively [5]. Favorable attitudes have not translated uniformly into clinical use: students and clinicians frequently report positive general views while hesitating to use AI in direct clinical work, citing accuracy, accountability, training, and the absence of institutional guidance [6–8]. The distance between access and responsible use makes AI preparedness an educational question as much as a technological one. It concerns how pre-registration programs and continuing professional development equip learners and practitioners to judge whether, when, and how AI output can inform professional work.

Technology acceptance research offers one way to frame that question. The Technology Acceptance Model (TAM) proposes that perceived usefulness (the degree to which a technology is believed to enhance performance) and perceived ease of use (the degree to which using it is expected to be effortless) are primary determinants of the intention to use it [9]. The Unified Theory of Acceptance and Use of Technology (UTAUT) adds social influence and facilitating conditions and treats experience as a moderator of several relationships [10]. Across healthcare technology acceptance research, the association between perceived usefulness and behavioral intention is the most frequently confirmed relationship [11], and meta-analytic evidence identifies performance expectancy, the UTAUT analogue of perceived usefulness, as the strongest single predictor of usage intention [2]. In medical education, TAM constructs have been used to model students' intentions to use ChatGPT and other large language models for learning [12,13]. Recent surveys of health professions students and rehabilitation professionals have likewise linked perceived usefulness to intention to use or actual use of generative AI [14,15], although among rehabilitation professionals, including SLPs, peer influence was the strongest predictor of use [14]. In both models, intention is the construct most directly linked to subsequent use, which makes it an informative outcome for education, although it is not equivalent to actual or appropriate use. The emphasis on core appraisals is not uncontested: reviewing TAM and UTAUT applications in healthcare, Lee et al. [1] argue that usefulness and ease-of-use appraisals are insufficient on their own in complex clinical environments and that trust, training, and organizational support require explicit modeling. The present study draws on this literature as a conceptual framework rather than testing either model, because its questionnaire does not measure every construct either model specifies.

A further question is whether direct experience with AI tools and knowledge about AI relate similarly to adoption intention. Among graduating medical students, previous experience with AI has been associated with more positive appraisals of its clinical value [4], whereas among health care students, self-perceived knowledge of ChatGPT and related language models was associated with more favorable attitudes toward those tools [8]. Neither study modeled professional adoption intention as an outcome or contrasted experiential engagement with self-rated knowledge. The distinction has curricular consequences, because knowledge-oriented AI literacy and supervised engagement with AI in professional tasks imply different educational designs.

Within speech-language pathology and therapy, AI applications span several clinical domains. Machine learning models have been applied to the automatic detection and classification of speech disorders, with dysarthria the most frequently studied condition [16], and to the analysis of connected speech in aphasia, where Cordella et al. [17] review exemplar applications alongside the reliability, validity, bias, and explainability concerns such methods raise. Generative tools have been proposed as aids for intervention planning, therapy material development, and documentation, which together constitute a substantial portion of clinical workload [7,18,19]. These tasks depend on clinical reasoning and individualized judgment, so the professional question is less whether AI can produce plausible output than

whether clinicians can recognize when that output is inaccurate, generic, or inappropriate for a particular client. Integration accordingly raises concerns about the clinical accuracy of AI-generated content, overreliance, data privacy and algorithmic bias, and the displacement of professional judgment [6,7,17]. Focusing on the preparation of student clinicians, Mytsyk and Suchikova [18] argue that responsible use requires explicit attention to misinformation, overreliance, and the critical evaluation of AI-generated output. Evaluations of generative AI output in the field remain small in scale: 15 practicing speech-language therapists rating ChatGPT-4.0 across six professional task domains reported high accuracy and relevance for report writing, clinical decision support, and educational material generation, but more variable ratings for therapy stimulus generation and therapy planning [20].

Research on AI perceptions within the profession has emerged mainly in the past three years. Aggarwal et al. [6] surveyed audiologists and speech-language pathologists and found that almost half viewed AI as helpful for diagnosis and treatment planning while identifying gaps in understanding and practical readiness. Surveying 106 academicians in the two professions internationally, Aggarwal et al. [21] documented wide variation in self-reported use alongside concerns about data authenticity, security, and citation accuracy. A scoping review of student attitudes across several health professions, including speech pathology, found broadly positive attitudes and willingness to learn, with negative perceptions concentrated on job security and on the perceived gap between AI capabilities and clinical realism [22]. More recent surveys have extended this work to 587 speech and language therapists in German-speaking countries, most of whom were open to the use of AI in the profession and whose most frequent concern was the loss of the "human factor" in assessment and therapy [23]; to 214 school-based SLPs in Ohio, United States, who reported using AI across a wide range of tasks while citing concerns about accuracy and ethics and a need for guidance and training [24]; and to 286 audiology and speech-language pathology students in India, most of whom used AI tools for educational purposes [25]. Like the earlier studies, these surveys were largely descriptive. The most directly relevant study, and the source of the present instrument, surveyed 107 speech-language pathology (SLP) clinicians and students in the United States [7]. Participants held broadly positive attitudes yet used AI infrequently for clinical purposes, restricting use largely to administrative tasks, and cited ethics, institutional guidance, and training as barriers.

In Türkiye, Birol et al. [20] and Yaşa et al. [26] evaluated AI-generated output against clinicians' judgments. Şimşek et al. [27] surveyed 168 undergraduate speech and language therapy students across class levels with a researcher-developed questionnaire and reported that ChatGPT was used mainly for academic purposes and less for clinical reasoning and case analysis, that students emphasized the need for institutional AI-use guidelines and AI literacy workshops, and that most regarded ChatGPT as a supportive tool that cannot replace the therapist.

These studies show that speech-language therapy students use generative AI and hold differentiated views of its academic and clinical value. They do not establish whether similar patterns hold among practicing clinicians, whether students and clinicians differ when assessed on identical dimensions, which AI-related appraisals are independently associated with intended use in professional practice rather than for learning, or whether experience with AI and self-rated knowledge are independently associated with intended use. Austin et al. [7] assessed clinicians and students with a common instrument but did not formally model factors associated with likelihood of AI use. These questions are relevant to the design of AI preparation in pre-registration programs and continuing professional development. The present study addresses these questions in three respects. First, it assesses final-year students and practicing clinicians in Türkiye with the same adapted instrument and describes, post hoc, the structure of the adapted scores in each group. Second, rather

than describing attitudes, it models which appraisals are independently associated with intended use in professional practice and examines the stability of these associations across alternative specifications. Third, it examines experiential engagement with AI alongside self-rated knowledge as correlates of intended use.

The present study therefore administered a Turkish adaptation of the questionnaire developed by Austin et al. [7] to speech-language therapy students and practicing clinicians. Its primary research question was: Which of self-rated AI knowledge, frequency of AI use, comfort with AI use, experience with ChatGPT/AI, concern regarding AI replacing speech-language therapy tasks, and perceived usefulness are independently associated with likelihood of AI use in speech-language therapy practice? Secondary questions concerned participants' perceptions across the questionnaire dimensions, differences between students and practicing clinicians, and associations among the dimensions. Optional open-ended responses were examined as an exploratory adjunct to contextualize the quantitative findings. Throughout, likelihood of AI use denotes the measured questionnaire dimension, intended use the construct it operationalizes, and behavioral intention the corresponding TAM construct.

## 2. Materials and Methods

### *2.1. Study Design and Setting*

This study used a cross-sectional online survey design to examine AI-related perceptions and intended AI use among speech-language therapy students and practicing clinicians in Türkiye. Cross-sectional surveys have been widely used to examine AI-related attitudes across healthcare and allied health professions [7,22]. Data were collected between 22 September and 10 October 2026. Because all data were collected at a single time point, the associations reported are correlational. The study is reported in accordance with the STROBE statement for cross-sectional studies [28]; the completed checklist is provided in File S1.

### *2.2. Participants*

The target population comprised speech-language therapists (SLTs) actively practicing in Türkiye and final-year (fourth-year) students enrolled in undergraduate speech and language therapy programs in Türkiye. In Türkiye, the profession is officially designated speech and language therapy, and practitioners hold the title speech-language therapist; these terms are used for participants and their professional context, whereas speech-language pathology is retained when referring to work from other national contexts. Final-year students were targeted because their coursework and clinical practicum exposure most closely approximate the professional knowledge and practice context relevant to the study's focus. Convenience sampling was used because a comprehensive, accessible national sampling frame was unavailable; the sample is therefore not assumed to be representative of students or clinicians nationally. The survey was disseminated through faculty advisors, class representatives, and student communication networks across all 31 speech and language therapy departments with fourth-year cohorts in Türkiye [29], and was also shared on social media. Practicing SLTs were reached through social media accounts and WhatsApp groups used for professional and academic exchange. The survey did not record institution or region, so the number of respondents from each department and the regional distribution of the sample are unknown. Eligibility criteria were (a) residence in Türkiye, (b) being a native Turkish speaker, and (c) current enrollment as a final-year speech and language therapy student or current practice as a qualified SLT. These criteria were stated in the information and consent text at the start of the survey, and participants confirmed that they met them before proceeding. Professional status was established by self-report on the educational-status item, and the other criteria were not verified independently. Professional

experience was recorded in four categories, the first three worded in Turkish as 0–2, 2–5, and 5–10 years "yıla kadar" (that is, until the upper value is reached) and the fourth as 10 years or more; the categories therefore corresponded to <2, 2 to <5, 5 to <10, and ≥10 years and did not overlap. To limit duplicate participation, the survey platform's setting allowing one response per device was enabled; in addition, IP addresses were screened, and 15 responses came from 7 shared addresses (11 in the analytic sample), none with identical response patterns, which is consistent with different respondents sharing a network.

A total of 443 individuals entered the survey. Of these, 66 (14.9%) did not provide the required response to an embedded instructed-response attention-check item asking participants to select the "Somewhat agree" option and were excluded from the analytic sample [30]; this exclusion criterion was established before data analysis. Reliance on a single embedded indicator is a recognized constraint, because one item cannot distinguish inattentive responding from isolated item-level error [30]. In a post hoc comparison, excluded and retained respondents did not differ significantly in professional status (students: 47.0% vs. 37.1%; $\chi^2$ [1] = 2.29, $p$ = 0.130), gender (female: 86.4% vs. 93.4%; Fisher's exact $p$ = 0.074; female vs. male or not stated), or age group (18–22, 23–27, and ≥28 years; $\chi^2$ [2] = 1.50, $p$ = 0.473). All 377 participants who passed the attention check provided complete responses to the closed-ended items, and no additional exclusions were applied. The final analytic sample comprised $N$ = 377 participants: 140 students and 237 practicing SLTs.

### *2.3. Instrument*

Data were collected using a Turkish adaptation of the multidimensional questionnaire developed by Austin et al. [7] for speech-language pathologists and students. Permission to translate the instrument into Turkish and to make the cultural and contextual modifications required for its use in Türkiye was obtained from the corresponding author of the source study before data collection (I. Ghanim, personal communication, 3 March 2026). Formal permission to translate and culturally adapt the questionnaire (Appendix A of Austin et al. [7]) was subsequently granted by the American Speech-Language-Hearing Association (ASHA), the copyright holder; the source material is copyrighted by ASHA and is used with permission. The questionnaire assesses conceptually distinct dimensions: self-rated knowledge of AI technologies, frequency of AI use, comfort with AI use, experience with ChatGPT/AI, concern regarding AI replacing SLT tasks, perceived accuracy of AI-generated assessment scores and assessment results, perceived client satisfaction with AI, perceived usefulness of AI in SLT practice, likelihood of AI use in SLT practice, and perceived effectiveness of AI for student skill development. The frequency and experience dimensions both concern AI use rather than evaluations of it: frequency indexes general habitual use, whereas experience indexes how often respondents had previously used AI in eight specified situations (e.g., under time pressure, to reduce workload, and in personal and professional life). Perceived usefulness and likelihood of AI use also differ in scope: the usefulness items ask how useful AI could be across six broad areas of professional work (use during therapy sessions; paperwork and administrative duties; other professional activities such as research, material development, and supervision; the education of future clinicians; building therapeutic relationships and trust with clients; and supporting clients' social skill development), whereas the likelihood items ask how likely respondents would be to use AI themselves for five specific clinical tasks (assessment, intervention, clinical evaluation and decision-making, therapy planning, and progress notes or report writing). Two pairs of areas are close in content: use during therapy sessions and intervention, and administrative duties and report writing. The two perceived-accuracy dimensions refer to successive steps of standardized assessment: the assessment-scores dimension asks whether current AI could correctly administer and score standardized tests in each clinical domain, whereas

the assessment-results dimension asks whether it could analyze the results of standardized assessment instruments to reach accurate conclusions. The contextual adaptations are summarized in Table S3.

The Turkish version retained the original item structure for most dimensions, including the three-item self-rated knowledge dimension, which comprises two items in which respondents appraise their own knowledge of AI and one item indexing incidental exposure to AI-related information; the dimension therefore indexes self-appraised familiarity rather than tested knowledge. Six modifications bear directly on the analyses. First, in the checklist of specific AI technologies, Grammarly, Amazon Alexa, and Google Project Euphonia were omitted as less contextually relevant for a Turkish sample; Google Gemini was added, and ChatGPT, AI image-generation tools, and Apple Siri were retained. Second, rather than averaging the five technology-specific frequency ratings of the source instrument, frequency of use was operationalized with two general items on habitual AI and ChatGPT use, to reduce dependence on the local availability of particular platforms. The source study's five-technology frequency composite had shown low internal consistency ($\alpha = 0.46$) [7], which is consistent with the possibility that averaging across platforms of differing local availability introduces heterogeneity; the frequency dimension is therefore a contextual modification rather than an item-for-item translation, and direct numerical comparison with the source study should be made cautiously. Third, three of the five task labels shared by the likelihood-of-use and replacement-concern items were adapted: lesson plans, a term tied to school-based service delivery, which is not a typical practice setting for speech and language therapists in Türkiye, was replaced by therapy planning, and evaluation and progress notes were specified as clinical evaluation and decision-making and as progress notes or report writing; because likelihood of AI use is the outcome variable in the primary regression, the outcome measure is not an item-for-item equivalent of the original. Fourth, one of the six perceived-usefulness items was expanded with clarifying examples of other professional contexts. Fifth, the two perceived-accuracy dimensions, which the source instrument administers over different clinical-domain lists, were administered over the same seven clinical domains, so the dimension derived from the eight-domain source list contains one item fewer (Table S3). Sixth, the client-satisfaction stem, which in the source asks how pleased clients would be to interact with AI during a therapy session, was rephrased to refer to AI used by the therapist as an auxiliary tool during therapy, because a literal translation was not readily comprehensible in Turkish.

Items were rated on five-point response scales with anchors varying by construct, and dimension scores were computed as item means, with higher scores indicating a higher level of the construct named by each dimension (for replacement concern, greater concern). The survey also included optional open-ended items on institutional practices and written policies regarding AI, possible challenges and benefits of AI use in therapy settings, and additional opinions or recommendations.

Because the instrument comprises conceptually distinct domains rather than indicators of a single latent construct, internal consistency estimates were examined as sample-specific checks of score coherence rather than as validation criteria. Within several dimensions, most clearly perceived usefulness, the items also refer to distinct areas of practice or tasks rather than restating a single attribute, so the dimension scores summarize appraisals across these areas and need not form a single factor. The study was not designed as a psychometric validation study, and factor-analytic validation was not undertaken as part of the original design; post hoc confirmatory factor and measurement invariance analyses are reported to characterize the structure of the adapted scores in this sample (Section 2.6). This approach follows the source study, which likewise computed construct scores by averaging items, reported internal consistency separately for each construct, and retained constructs with

low coefficients in its analyses without factor-analytic evaluation [7]. With the permission of the copyright holder (ASHA), the administered Turkish questionnaire was provided to the editor and reviewers as a confidential file for peer review, but it is not publicly available or published with the article. The adaptations to the questionnaire dimensions are summarized in Table S3. Scoring followed the source study, and evidence on the structure of the outcome measure and on the distinctness of perceived usefulness and likelihood of AI use is reported in Sections 3.2 and 3.5.

### 2.4. Adaptation Procedure

The adaptation was informed by established cross-cultural adaptation guidance [31] and proceeded after permission had been obtained from the source study's corresponding author. Two researchers independently translated the questionnaire into Turkish, attending to linguistic equivalence and conceptual-cultural appropriateness, and reconciled their versions through consensus discussion. Two speech and language therapy academics then reviewed the unified version for semantic equivalence, disciplinary terminology, and contextual relevance, and the context-specific modifications described above were introduced at this stage. An independent bilingual researcher, unaware of the original instrument, back-translated the revised version into English, and comparison confirmed that the conceptual meaning of the retained content had been preserved. A small pilot group of speech and language therapy students reviewed the Turkish version for clarity, and minor wording refinements were made before data collection. The process was designed to support linguistic and contextual appropriateness rather than to establish a newly validated Turkish scale, and psychometric findings are interpreted as sample-specific properties of the adapted questionnaire.

### 2.5. Data Collection

Data were collected via the SurveyMonkey platform between 22 September and 10 October 2026. The survey link was disseminated through professional networks, social media, and university communication channels targeting speech and language therapy students and clinicians in Türkiye. All participants provided electronic informed consent before participation and were informed that participation was voluntary and anonymous. No names or contact details were collected; IP addresses recorded automatically by the survey platform were used only to screen for duplicate responses and were deleted from the dataset after screening. Estimated completion time was 10–15 min. Ethics approval is reported in the Institutional Review Board Statement.

### 2.6. Data Analysis

All analyses were performed in jamovi (version 2.7.38) [32]. Before analysis, the dataset was screened for missing data and attention-check failures; no responses were missing on the closed-ended items, and all dimension scores were complete for the analytic sample. Descriptive statistics (frequencies, percentages, means [$M$], standard deviations [$SD$], medians [$Mdn$], and ranges) were computed for all demographic and questionnaire variables, and internal consistency was evaluated with Cronbach's α for each dimension as a descriptive index of within-dimension homogeneity.

Because Shapiro–Wilk tests indicated significant departures from normality in several dimensions, nonparametric procedures were used for bivariate analyses. Spearman's rank-order correlations (ρ) were computed for all 55 pairwise combinations of the 11 dimensions, with the Benjamini–Hochberg procedure applied at $q = 0.05$ to control the false discovery rate across this exploratory family. Differences between students and practicing SLTs were examined with Mann–Whitney $U$ tests, with effect sizes indexed by the rank-biserial

correlation ($r_{rb}$); because $r_{rb}$ and the Pearson correlation are not on the same metric, Cohen's benchmarks for $r$ (small ≈0.10, medium ≈0.30, large ≈0.50) [33] are used only as a coarse interpretive frame. A Bonferroni-corrected threshold ($\alpha = 0.0045$) was applied to the 11 group comparisons. False discovery rate control was used for the larger exploratory correlation family, and the more conservative Bonferroni correction for the smaller family of student–clinician comparisons.

A simultaneous multiple linear regression was conducted to identify variables independently associated with likelihood of AI use. Six variables were entered in a single block: self-rated knowledge of AI technologies, frequency of AI use, comfort with AI use, experience with ChatGPT/AI, concern regarding AI replacing SLT tasks, and perceived usefulness of AI in practice. These six were selected on conceptual grounds because each corresponds to a construct discussed as a potential correlate of adoption intention in the TAM and UTAUT literature: self-rated knowledge, habitual use, comfort, experiential familiarity, perceived professional threat, and perceived usefulness. Perceived usefulness is itself an appraisal; it was retained because it corresponds directly to the core TAM construct concerning the value of AI for professional work in general. The four remaining dimensions concern judgments of AI performance on specific assessment, client-response, and training outcomes, and were not entered in the primary model because their position relative to perceived usefulness and intended use could not be specified from the framework. The model specification was not preregistered.

Assumptions were evaluated following standard recommendations for multiple regression [34] through residual plots, Cook's distance, Mahalanobis distance, and variance inflation factors (VIFs): the residual Shapiro–Wilk test indicated a small departure from normality ($W = 0.991$, $p = 0.025$), Cook's distances were uniformly low (maximum = 0.054), two observations flagged by the Mahalanobis criterion showed no undue influence and were retained, and VIFs ranged from 1.03 to 1.95. Ninety-five percent confidence intervals (CIs) are reported for all coefficients, and significance was set at $p < 0.05$. Given the small departure from residual normality, estimate stability was evaluated with nonparametric bootstrap resampling (5000 replications), yielding percentile 95% CIs for the unstandardized coefficients. Because the design is cross-sectional, coefficients index independent association among concurrently measured variables rather than causal or temporal effects.

Two post hoc sensitivity analyses re-estimated the model (a) with a two-item knowledge score excluding the exposure item and (b) with the four evaluative dimensions entered alongside the primary variables (Tables S1 and S2). Coefficient-level inference in these models used HC3 heteroskedasticity-consistent standard errors, whereas the primary model uses conventional ordinary least squares (OLS) standard errors with bootstrap intervals, so $p$-values are not directly comparable across the primary and sensitivity models. To describe how the explained variance was distributed among the predictors, a post hoc general dominance analysis [35], run in jamovi with the Rj editor, additionally decomposed the model $R^2$ into the average contribution of each variable across all subset models, with percentile bootstrap 95% CIs (1000 replications) for each contribution and for the difference between the experience and self-rated knowledge contributions (Table S8). No a priori sample size calculation was conducted. Further post hoc sensitivity analyses re-estimated the primary model (c) including the 66 respondents who failed the attention check, (d) retaining only the first response from each IP address, and (e) replacing the perceived-usefulness score with two subscores for areas outside the therapeutic encounter (administrative work, other professional activities, and the education of future clinicians) and within it (use during sessions, therapeutic relationships, and clients' social skills) (Table S12). To characterize the structure of the adapted scores, a single-factor confirmatory factor analysis was estimated post hoc for each dimension with four or more items, using robust maximum likelihood,

which simulation evidence supports for items with five or more response categories, although it can be biased when category thresholds are asymmetric [36]; McDonald's $\omega$ and corrected item–total correlations were also computed (Table S10). For the same dimensions, measurement invariance between students and practicing SLTs was examined with multigroup models testing configural, metric (equal loadings), and scalar (equal loadings and intercepts) invariance, judged by the scaled $\chi^2$ difference test and by changes in CFI and RMSEA, with a decrease in CFI of 0.010 or more or an increase in RMSEA of 0.015 or more taken to indicate noninvariance [37] (Table S11). Because invariance tests presuppose an adequately fitting configural model, changes in fit under equality constraints were not interpreted as evidence of invariance when the configural model fitted poorly. For the dimensions whose single-factor models fitted poorly and whose local misfit was concentrated in item pairs sharing identifiable content (perceived usefulness, experience with ChatGPT/AI, and perceived client satisfaction), the models were respecified post hoc by allowing residual covariances for those pairs, and the respecified models were re-estimated with weighted least squares mean- and variance-adjusted (WLSMV) estimation treating items as ordinal, as a sensitivity check (Tables S10 and S11). These analyses were conducted in R with lavaan (version 0.6-17) [38]. Further post hoc analyses examined (f) the three knowledge items individually, as bivariate correlates of likelihood of AI use and entered separately in the primary model (Table S13); (g) the distinctness of perceived usefulness and likelihood of AI use, by comparing one-factor and two-factor confirmatory models of their items, estimated in the same way, with absolute fit indices and the latent correlation reported for each model; (h) the primary model with perceived usefulness scored without the two items closest in content to likelihood-of-use items (use during sessions and administrative work); (i) the primary model with each likelihood-of-use task as a separate outcome (Table S14); and (j) the primary model as a structural equation model in which perceived usefulness and likelihood of AI use were specified as latent variables (Table S12).

Students and practicing SLTs were combined for the internal consistency estimates, the correlational analyses, and the regression model, to preserve power for the multivariable model. Because non-significant group differences would not establish equivalence, pooled estimates are interpreted as describing the combined sample. As sensitivity analyses, professional status (student vs. practicing SLT) was added to the regression model as a covariate, and the model was re-estimated separately in each group. Post hoc, the six professional status × variable interaction terms were tested jointly with an $F$ test, and individual interaction terms were evaluated against a Bonferroni-corrected threshold ($\alpha = 0.0083$). In the clinician subsample, the model was also re-estimated with years of professional experience and work setting (special education and rehabilitation center vs. other) as covariates.

The first author conducted the initial inductive coding of the open-ended responses; the second and third authors independently reviewed the coding structure and thematic categorization, and discrepancies were discussed until consensus was reached. The procedure is best described as inductive, content-oriented thematic coding of brief open-ended survey responses; it was not conducted as reflexive thematic analysis. The unit of coding was the idea unit, so a single response could contribute to more than one category, and categories are reported descriptively rather than as frequency counts. Because the review was consensus-based verification rather than independent duplicate coding, an inter-coder reliability coefficient was not calculated. Most participants responded to each open-ended item (students $n$ = 130–132; clinicians $n$ = 221–233; Table S7), but responses were typically brief. Responses were written in Turkish; quotations were translated into English, with the Turkish originals in Table S4. Participant identifiers (ST = student, CL = clinician) were assigned for reporting only.

### 2.7. Use of Large Language Models

Claude (Anthropic) was used solely for English language editing, translation, and improving the linguistic clarity and readability of the manuscript. The tool was not used for data collection, data analysis, statistical calculations, interpretation of findings, or generation of scientific content. All authors reviewed and approved the final manuscript and take full responsibility for its content.

## 3. Results

### 3.1. Participant Characteristics

Demographic and AI-familiarity characteristics of the sample are summarized in Table 1. Both groups were predominantly female (students: 90.0%; practicing SLTs: 95.4%), and age distributions reflected the groups' differing career stages, with students concentrated in the 18–27 range. Among practicing SLTs, most worked in special education and rehabilitation centers (68.8%), held a bachelor's degree (89.5%), and reported <2 years of professional experience (69.6%); the practicing sample was therefore predominantly early-career.

**Table 1.** Demographic characteristics and AI-related familiarity of the sample ($N = 377$).

| Variable | Students, *n* (%) | Practicing SLTs, *n* (%) |
|---|---|---|
| **Gender** | | |
| Female | 126 (90.0) | 226 (95.4) |
| Male | 13 (9.3) | 10 (4.2) |
| Prefer not to say | 1 (0.7) | 1 (0.4) |
| **Age group (years)** | | |
| 18–22 | 80 (57.1) | 8 (3.4) |
| 23–27 | 53 (37.9) | 196 (82.7) |
| 28–32 | — | 24 (10.1) |
| 33–37 | 1 (0.7) | 4 (1.7) |
| 38–42 | 2 (1.4) | 4 (1.7) |
| ≥43 | 4 (2.9) | 1 (0.4) |
| **Primary work setting (practicing SLTs only)** | | |
| Special education/rehabilitation center | — | 163 (68.8) |
| Private clinic | — | 45 (19.0) |
| Hospital or medical setting | — | 10 (4.2) |
| University | — | 5 (2.1) |
| Municipal rehabilitation center | — | 3 (1.3) |
| Other | — | 11 (4.6) |
| **Highest completed degree (practicing SLTs only)** | | |
| Bachelor's degree | — | 212 (89.5) |
| Master's degree | — | 23 (9.7) |
| Doctoral degree | — | 2 (0.8) |
| **Years of professional experience (practicing SLTs only)** | | |
| <2 years | — | 165 (69.6) |
| 2 to <5 years | — | 49 (20.7) |
| 5 to <10 years | — | 19 (8.0) |
| ≥10 years | — | 4 (1.7) |
| **AI tools recognized** | | |
| ChatGPT | 137 (97.9) | 233 (98.3) |
| Gemini | 131 (93.6) | 196 (82.7) |
| Apple Siri | 98 (70.0) | 137 (57.8) |
| AI image generation tools | 67 (47.9) | 83 (35.0) |
| Other [a] | 7 (5.0) | 19 (8.0) |

**Table 1.** *Cont.*

| Variable | Students, *n* (%) | Practicing SLTs, *n* (%) |
|---|---|---|
| **Frequency of AI use** | | |
| Never | — | 4 (1.7) |
| Less than once per week | 4 (2.9) | 13 (5.5) |
| Approximately 1–2 times/week | 35 (25.0) | 41 (17.3) |
| Approximately 3–5 times/week | 50 (35.7) | 87 (36.7) |
| More than once daily | 51 (36.4) | 92 (38.8) |
| **General attitudes toward AI and ChatGPT** | | |
| Extremely negative | — | 1 (0.4) |
| Somewhat negative | 3 (2.1) | 11 (4.6) |
| Neither negative nor positive | 20 (14.3) | 28 (11.8) |
| Somewhat positive | 69 (49.3) | 121 (51.1) |
| Extremely positive | 48 (34.3) | 76 (32.1) |
| **Total** | 140 (100.0) | 237 (100.0) |

Percentages are reported separately for each group. Experience categories follow the survey wording, in which each of the first three categories extended until, but did not include, its upper value. Participants could select multiple AI tools, so column totals in the AI tools section exceed 100%. AI = artificial intelligence; SLT = speech-language therapist.[a] Includes Microsoft Copilot, Grok, Yandex AI, Learna AI, NotebookLM, Claude, DeepSeek, Anara, Consensus, Canva AI, Suno, Skywork.ai, and Kumru.

ChatGPT was the most widely recognized AI tool in both groups (97.9% of students and 98.3% of practicing SLTs), with Gemini and Apple Siri the next most commonly identified. Most respondents in both groups reported using AI at least three to five times per week, and approximately 83% of each group reported somewhat or extremely positive attitudes toward AI.

### *3.2. Descriptive Statistics and Internal Consistency*

Table 2 presents descriptive statistics and Cronbach's α for each dimension in the combined sample ($N = 377$). Frequency of AI use obtained the highest mean score ($M = 4.03$). Mean scores for perceived accuracy of assessment scores ($M = 2.44$), perceived client satisfaction ($M = 2.51$), perceived accuracy of assessment results ($M = 2.72$), and effectiveness of AI for student skill development ($M = 2.77$) fell below the scale midpoint, and replacement concern had the lowest mean ($M = 2.38$).

**Table 2.** Descriptive statistics and internal consistency estimates ($N = 377$).

| Questionnaire Dimension | *k* | α | *M* | *SD* | *Mdn* | Possible Range |
|---|---|---|---|---|---|---|
| 1. Self-rated knowledge of AI technologies | 3 | 0.497 | 3.42 | 0.56 | 3.33 | 1–5 |
| 2. Frequency of AI use | 2 | 0.935 | 4.03 | 0.91 | 4.00 | 1–5 |
| 3. Comfort with AI use | 8 | 0.866 | 3.20 | 0.68 | 3.12 | 1–5 |
| 4. Experience with ChatGPT/AI | 8 | 0.912 | 3.31 | 0.92 | 3.25 | 1–5 |
| 5. Concern: AI replacing SLT tasks | 5 | 0.859 | 2.38 | 0.86 | 2.20 | 1–5 |
| 6. Likelihood of AI use in SLT practice | 5 | 0.875 | 3.11 | 0.90 | 3.20 | 1–5 |
| 7. Perceived accuracy: assessment scores | 7 | 0.863 | 2.44 | 0.75 | 2.43 | 1–5 |
| 8. Perceived accuracy: assessment results | 7 | 0.917 | 2.72 | 0.83 | 2.86 | 1–5 |
| 9. Perceived client satisfaction with AI | 8 | 0.953 | 2.51 | 1.04 | 2.62 | 1–5 |
| 10. Perceived usefulness of AI in SLT practice | 6 | 0.760 | 3.08 | 0.68 | 3.00 | 1–5 |
| 11. Effectiveness of AI: student skill development | 9 | 0.917 | 2.77 | 0.85 | 2.89 | 1–5 |

*k* = number of items contributing to each dimension score; α = Cronbach's alpha; *M* = mean; *SD* = standard deviation; *Mdn* = median. All items used five-point response scales with anchors varying by dimension; higher scores indicate a higher level of the construct named by each dimension (for replacement concern, greater concern). α values are descriptive indices of within-dimension homogeneity and are not reported as scale-validation criteria; for dimensions 2 and 4, high α reflects consistent responding across related items and does not imply unidimensionality. AI = artificial intelligence; SLT = speech and language therapy.

Internal consistency was acceptable to excellent for 10 of the 11 dimensions ($\alpha = 0.760–0.953$). The two-item frequency dimension yielded $\alpha = 0.935$, implying an inter-item correlation of approximately 0.88 and indicating that the two items function as near-duplicates rather than complementary indicators. Perceived usefulness showed the lowest $\alpha$ among these ten dimensions (0.760). The three-item self-rated knowledge dimension showed markedly lower consistency ($\alpha = 0.497$) than the corresponding dimension in the source study ($\alpha = 0.76$) [7]. Across the remaining dimensions the two samples yielded closely comparable coefficients (e.g., experience, 0.912 vs. 0.92; replacement concern, 0.859 vs. 0.86; client satisfaction, 0.953 vs. 0.96), with two exceptions in opposite directions: the adapted frequency dimension was considerably more homogeneous than in the source study (0.935 vs. 0.46), whereas perceived usefulness was somewhat less so (0.760 vs. 0.87); one usefulness item had been expanded with clarifying examples in the adaptation (Table S3). The divergence is therefore dimension-specific and may partly reflect the heterogeneous, content-sampling nature of the items, the adaptation, or sample characteristics; these possibilities were not tested. Item-level diagnostics showed that the daily AI/ChatGPT information-exposure item had the weakest corrected item–total correlation in the knowledge dimension ($r = 0.178$); deleting it increased $\alpha$ to 0.664, but the three-item structure of the source questionnaire was retained. Internal consistency was similar in students and practicing SLTs for every dimension (largest difference in $\alpha$, 0.060 for self-rated knowledge; Table S10). For the nine dimensions with four or more items, $\omega$ ranged from 0.74 to 0.95 and corrected item–total correlations ranged from 0.37 to 0.87. A single-factor model fitted the likelihood-of-use items well according to CFI and SRMR, although RMSEA exceeded conventional cutoffs (CFI = 0.985, SRMR = 0.026, RMSEA = 0.086), whereas the other eight dimensions fitted poorly (CFI = 0.72–0.92, RMSEA = 0.14–0.27), indicating that these scores combine related but not strictly unidimensional ratings; the perceived-usefulness items showed the weakest single-factor structure (CFI = 0.72; loadings 0.28–0.83), with residual dependence between the administrative and other professional items and between the two items concerning clients' relationships and social skills. Allowing residual covariances for these two pairs improved fit (CFI = 0.956, TLI = 0.906, SRMR = 0.051), although RMSEA remained high (0.105, 90% CI [0.071, 0.142]), and the administrative and other professional items loaded weakly on the common factor (0.36 and 0.40; $\omega = 0.64$). Corresponding content-based respecifications improved fit for experience with ChatGPT/AI (CFI = 0.955, RMSEA = 0.112, SRMR = 0.038) and perceived client satisfaction (CFI = 0.947, RMSEA = 0.149, SRMR = 0.041), but RMSEA remained above conventional cutoffs (Table S10). The dimension scores are therefore treated as composites of related ratings rather than as measures of validated unidimensional constructs. In multigroup models, the configural model fitted acceptably only for likelihood of AI use (CFI = 0.988, RMSEA = 0.078), for which equality constraints on loadings and intercepts did not worsen fit (scalar $\Delta$CFI = 0.001, $p = 0.599$). For the other eight dimensions, the configural models fitted poorly, including after respecification (perceived client satisfaction, CFI = 0.939, RMSEA = 0.160; perceived usefulness, CFI = 0.957, RMSEA = 0.104), so the changes in fit under equality constraints were not interpreted as evidence of metric or scalar invariance (Table S11). Comparisons between students and clinicians on these dimensions therefore concern observed scores. The three knowledge items correlated weakly with each other ($r = 0.11–0.51$), and only the knowledge-rating and exposure items were associated with likelihood of AI use at the bivariate level (both $\rho = 0.18$, $p < 0.001$; comfort explaining AI, $\rho = 0.09$, $p = 0.081$). Entered separately in the primary model, none of the three items was positively and independently associated with the outcome, and the item on comfort explaining AI showed a small negative coefficient ($\beta = -0.100$, $p = 0.046$) (Table S13).

### *3.3. Correlations Among Dimensions*

Spearman correlations among all dimensions are reported in Table S5. After Benjamini–Hochberg correction across the 55 tests, one small association between self-rated knowledge and perceived effectiveness of AI for student skill development ($\rho = 0.10$, unadjusted $p = 0.046$), was no longer significant (adjusted $p = 0.056$); all other associations significant at the unadjusted level remained so.

Frequency of use, comfort, and experience with ChatGPT/AI were positively intercorrelated: experience correlated with comfort and with frequency of use (both $\rho = 0.53$, $p$s $< 0.001$), and self-rated knowledge correlated somewhat less strongly with comfort ($\rho = 0.44$), frequency ($\rho = 0.40$), and experience ($\rho = 0.40$; all $p$s $< 0.001$).

Likelihood of AI use was most strongly associated with perceived usefulness ($\rho = 0.52$, $p < 0.001$), followed by perceived client satisfaction ($\rho = 0.42$), experience with ChatGPT/AI ($\rho = 0.41$), perceived effectiveness of AI for student skill development ($\rho = 0.38$), perceived accuracy of assessment results ($\rho = 0.35$), and perceived accuracy of assessment scores and comfort with AI use (both $\rho = 0.34$; all $p$s $< 0.001$). Self-rated knowledge was more weakly associated with likelihood of AI use ($\rho = 0.21$, $p < 0.001$).

The two perceived-accuracy dimensions were strongly intercorrelated ($\rho = 0.67$, $p < 0.001$). Replacement concern showed weak and largely non-significant associations with most dimensions; its significant associations were with likelihood of AI use ($\rho = 0.19$) and perceived accuracy of assessment scores ($\rho = 0.19$; both $p$s $< 0.001$), perceived accuracy of assessment results ($\rho = 0.14$, $p = 0.008$), and experience with ChatGPT/AI ($\rho = 0.13$, $p = 0.009$), all of which survived correction. Replacement concern was therefore at most weakly associated with the other dimensions.

### *3.4. Comparison of Questionnaire Scores Between Students and Practicing Clinicians*

Mann–Whitney $U$ test results are reported in Table S6. After application of the Bonferroni-corrected threshold ($\alpha = 0.0045$), only perceived client satisfaction differed significantly between groups. Unadjusted differences reached the conventional 0.05 level for replacement concern ($p = 0.006$) and self-rated knowledge ($p = 0.046$) but did not survive correction, and the remaining eight dimensions (likelihood of AI use, perceived usefulness, frequency and comfort of use, experience with ChatGPT/AI, both perceived-accuracy dimensions, and effectiveness of AI for student skill development) differed at neither threshold (all $p$s $\geq 0.075$).

Practicing SLTs reported more positive perceptions of client satisfaction with AI than students ($U = 13{,}382$, $p = 0.002$, $r_{rb} = 0.19$). Students reported higher replacement concern ($U = 13{,}786$, $p = 0.006$, $r_{rb} = -0.17$) and higher self-rated knowledge of AI technologies ($U = 14{,}586$, $p = 0.046$, $r_{rb} = -0.12$); both are reported descriptively. All effects were small. Because measurement invariance was not established for these dimensions (Section 3.2), all group comparisons concern observed scores.

### *3.5. Factors Associated with Likelihood of AI Use*

The six-variable model was statistically significant, $F(6, 370) = 34.48$, $p < 0.001$, accounting for 35.9% of the variance in likelihood of AI use ($R^2 = 0.359$, adjusted $R^2 = 0.348$), a large effect by conventional benchmarks ($f^2 = 0.56$) [33] (Table 3). Perceived usefulness of AI in practice showed the largest independent association with likelihood of AI use ($\beta = 0.427$, $p < 0.001$), followed by experience with ChatGPT/AI ($\beta = 0.239$, $p < 0.001$). Concern regarding AI replacing SLT tasks was also positively, though weakly, associated with likelihood of AI use ($\beta = 0.110$, $p = 0.010$). Self-rated knowledge, frequency of AI use, and comfort with AI use were not significantly associated with likelihood of AI use after adjustment for the other variables. VIFs ranged from 1.03 to 1.95.

**Table 3.** Multiple linear regression model for likelihood of AI use in SLT practice ($N = 377$).

| Variable | *B* | *SE* | β | 95% CI for *B* | *t* | *p* |
|---|---|---|---|---|---|---|
| Constant | 0.057 | 0.293 | — | [−0.519, 0.633] | 0.195 | 0.846 |
| Self-rated knowledge of AI technologies | −0.010 | 0.080 | −0.006 | [−0.167, 0.148] | −0.121 | 0.904 |
| Frequency of AI use | 0.035 | 0.052 | 0.035 | [−0.067, 0.137] | 0.674 | 0.501 |
| Comfort with AI use | 0.043 | 0.072 | 0.032 | [−0.100, 0.185] | 0.588 | 0.557 |
| Experience with ChatGPT/AI | 0.235 | 0.057 | 0.239 | [0.122, 0.347] | 4.101 | <0.001 |
| Concern: AI replacing SLT tasks | 0.116 | 0.045 | 0.110 | [0.028, 0.203] | 2.599 | 0.010 |
| Perceived usefulness of AI in SLT practice | 0.571 | 0.061 | 0.427 | [0.452, 0.691] | 9.417 | <0.001 |

$R^2 = 0.359$, adjusted $R^2 = 0.348$, $F(6, 370) = 34.48$, $p < 0.001$. β = standardized regression coefficient; CI = confidence interval. Values are conventional ordinary least squares estimates. Nonparametric percentile bootstrap 95% CIs (5000 replications) for the unstandardized coefficients were: self-rated knowledge [−0.168, 0.148], frequency [−0.066, 0.140], comfort [−0.109, 0.194], experience [0.120, 0.351], replacement concern [0.022, 0.214], and perceived usefulness [0.443, 0.691]. The professional status interactions and group-specific models reported in the text were post hoc analyses and are not part of the model shown (Table S9). AI = artificial intelligence; SLT = speech and language therapy.

Re-estimating the coefficients with 5000 bootstrap samples yielded the same pattern: percentile intervals for perceived usefulness [0.443, 0.691], experience with ChatGPT/AI [0.120, 0.351], and replacement concern [0.022, 0.214] excluded zero, whereas those for self-rated knowledge [−0.168, 0.148], frequency of use [−0.066, 0.140], and comfort [−0.109, 0.194] included it. Re-estimating the primary model with HC3 heteroskedasticity-consistent standard errors left the pattern of significance unchanged (experience, $p < 0.001$; replacement concern, $p = 0.020$; perceived usefulness, $p < 0.001$; all other $p$s $\geq 0.50$).

In the general dominance analysis, perceived usefulness accounted for 53.8% of the explained variance (contribution to $R^2 = 0.193$, 95% CI [0.125, 0.266]) and experience with ChatGPT/AI for 21.6% (0.077 [0.042, 0.124]), whereas self-rated knowledge accounted for 3.5% (0.013 [0.005, 0.033]); the difference between the experience and knowledge contributions excluded zero (0.065, 95% CI [0.025, 0.110]; Table S8).

When professional status was added to the model, it was not independently associated with likelihood of AI use (β = −0.004, $p = 0.932$), $R^2$ was unchanged, and the coefficient pattern was essentially unchanged. The six status × variable interaction terms did not jointly improve the model ($\Delta R^2 = 0.015$, $F$ [6, 363] = 1.47, $p = 0.186$). The interaction terms for experience with ChatGPT/AI ($p = 0.033$) and replacement concern ($p = 0.044$) reached the unadjusted 0.05 level but not the Bonferroni-corrected threshold. Exploratory group-specific models nevertheless showed a divergent pattern (Table S9). Perceived usefulness was independently associated with likelihood of AI use in both groups (students, β = 0.466; practicing SLTs, β = 0.402; both $p$s < 0.001). Experience with ChatGPT/AI was associated with the outcome among students (β = 0.409, $p < 0.001$) but not among practicing SLTs (β = 0.135, $p = 0.080$; 95% CI for $B$ [−0.016, 0.287]), whereas replacement concern was associated with the outcome among practicing SLTs (β = 0.177, $p = 0.002$) but not among students (β = 0.012, $p = 0.849$). In the clinician subsample, adding years of professional experience and work setting raised $R^2$ from 0.323 to 0.335; neither covariate reached significance, and the coefficient pattern was unchanged.

In the post hoc sensitivity analyses, re-estimating the model with a two-item knowledge score left the pattern unchanged (Table S1). Entering the four evaluative dimensions increased explained variance ($\Delta R^2 = 0.041$, $F$-change [4, 366] = 6.19, $p < 0.001$) and identified perceived client satisfaction as an additional independent correlate (β = 0.173, $p < 0.001$), whereas the two accuracy dimensions and the student-skill composite were not (Table S2). Variance inflation factors in this model ranged from 1.08 to 2.03. The two perceived-accuracy dimensions, which were strongly intercorrelated (ρ = 0.67), jointly

added little explained variance ($\Delta R^2 = 0.008$, $F$ [2, 366] = 2.45, $p = 0.088$; HC3-based Wald test, $p = 0.155$), so their individual coefficients are not interpreted separately. The perceived usefulness coefficient decreased from β = 0.427 to β = 0.276, indicating variance shared with the added dimensions; its magnitude was therefore sensitive to model specification. The replacement-concern coefficient also decreased ($B = 0.116$ to $B = 0.100$; β = 0.095), and its HC3-based $p$-value (0.048) lay close to the conventional threshold; because the primary model used conventional standard errors, this change reflects both the change in specification and the change in inferential method. Including the 66 respondents who failed the attention check ($N = 443$) left the associations with perceived usefulness (β = 0.404) and experience (β = 0.223; both $p < 0.001$) essentially unchanged, whereas that with replacement concern was smaller and no longer significant (β = 0.076, $p = 0.059$); $R^2$ was 0.331. Retaining only the first response from each IP address ($n = 371$) left the pattern unchanged ($R^2 = 0.359$). When perceived usefulness was split into subscores for areas outside and within the therapeutic encounter, both were independently associated with likelihood of AI use (β = 0.221 and β = 0.272, both $p < 0.001$), and $R^2$ was unchanged (0.359) (Table S12). In a confirmatory comparison of their items, neither a one-factor model (CFI = 0.735, RMSEA = 0.167, SRMR = 0.110) nor a two-factor model separating perceived usefulness from likelihood of AI use (CFI = 0.857, RMSEA = 0.124, SRMR = 0.092) fitted adequately, although the two-factor model fitted better (scaled $\Delta\chi^2$ [1] = 93.73, $p < 0.001$). With the two residual covariances among usefulness items allowed, the two-factor model fitted acceptably (CFI = 0.958, TLI = 0.944, RMSEA = 0.069, 90% CI [0.052, 0.085], SRMR = 0.051), and the latent correlation between the factors was 0.65 (95% CI [0.55, 0.76]); a residual correlation between the administrative-duties and report-writing items (0.19, 95% CI [0.10, 0.29]) indicated content overlap (Table S14). In a structural equation model in which perceived usefulness and likelihood of AI use were latent variables, the coefficient pattern was unchanged (perceived usefulness, β = 0.55, 95% CI [0.43, 0.68]; experience with ChatGPT/AI, β = 0.26, $p < 0.001$; replacement concern, β = 0.14, $p = 0.005$; other predictors nonsignificant), although coefficient magnitudes depended on the measurement model (Table S12). Scoring perceived usefulness without the two items closest in content to likelihood-of-use items reduced its coefficient from β = 0.427 to β = 0.355 ($p < 0.001$), with experience unchanged (β = 0.250). With each task as a separate outcome, perceived usefulness was independently associated with likelihood of use for all five tasks (β = 0.31–0.40, all $p < 0.001$), whereas experience was associated with four tasks (β = 0.16–0.34) but not with progress notes or report writing (β = 0.05, $p = 0.444$); intended use was highest for progress notes or report writing (62.6% would probably or definitely use AI) and therapy planning (54.1%) and lowest for clinical evaluation and decision-making (34.7%) and assessment (36.1%) (Table S14).

### *3.6. Open-Ended Responses*

Open-ended responses were analyzed as an exploratory element embedded within the survey rather than as a stand-alone qualitative study, and are reported as contextual support for the quantitative results. The categories identified in student and clinician responses are summarized in Table S7; illustrative responses are quoted below.

Both students and practicing clinicians viewed AI and ChatGPT primarily as supportive tools for material preparation, information access, and workload reduction rather than as replacements for clinical expertise or therapeutic interaction, while also raising concerns about misinformation, ethical responsibility, overreliance and reduced clinical reasoning, and the weakening of individualized, human-centered practice.

Responses to the item on possible challenges centered on the accuracy of AI-generated content and on its effect on the therapeutic relationship, and some responses linked the two.

One student described in-session use as something that could lead a client to question the purpose of attending therapy: "...asking ChatGPT in front of the patient and doing what it says can give the patient the feeling of 'then why am I here, I could look it up myself'" (ST1). A clinician framed the same tension as a competing demand on attention within the session: "The therapist may be thought to lack knowledge; within the same session it may be difficult both to maintain the therapeutic relationship and to use AI" (CL1). A second concern, expressed in both groups, was that AI output would be too generic for individualized planning: "It may misdirect scoring on standardized tests and may not perform detailed analysis in individualized therapy planning; the likelihood of remaining superficial is high" (CL2). A third concern, raised in both groups, addressed the clinician's own reasoning rather than the client: "The therapist's reasoning skills and capacity to think things through weaken" (CL3).

Perceived benefits clustered on preparatory and administrative work rather than on the therapeutic encounter itself, and could be named together in a single response: "Speeding up reporting work, doing the drudge work, generating ideas for therapy materials" (ST2); "Lightening the therapist's load, above all paperwork and report writing; offering information about speech and language skills in rarely seen syndromes; offering different therapy methods" (CL4). Endorsements could be bounded within the same sentence that expressed them. The clinician who warned that AI planning risks "remaining superficial" (CL2, above) also identified that same superficiality as a usable starting point: "it can produce a superficial draft for therapy planning, and therapies can then be enriched from that draft."

In the final open-ended item, a position expressed in both groups was conditional acceptance: AI as a supportive instrument whose value depends on the clinician's own judgment. "In my view AI and GPT are useful as a supportive, informative, awareness-raising tool, but they do not fully take a therapist's place" (CL5). In some responses, the conditions were stated explicitly: "I think it can be useful if it is not followed blindly, if it is used by writing detailed prompts, if the therapist personalizes it and blends it with their own knowledge, and if the accuracy of the information is checked" (ST3). Responses of this form, in which endorsement was conditional on verification, personalization, and professional oversight, appeared in both groups. In this item, students also referred to the need for controlled and supervised use, and clinicians to AI literacy and training needs.

In both groups, responses to the two items on institutional context described the absence of a written institutional AI policy or uncertainty about whether one existed. Students additionally described supportive educational use as well as academic monitoring or restrictions (Table S7).

## 4. Discussion

### *4.1. Main Findings*

This study examined which AI-related appraisals were independently associated with intended AI use among speech-language therapy students and practicing clinicians assessed with the same questionnaire. Perceived usefulness of AI in practice and experience with ChatGPT/AI showed the strongest independent associations with likelihood of AI use, in a model that accounted for 35.9% of its variance. Because several dimension scores, including perceived usefulness, showed weak single-factor structure (Section 3.2), these coefficients describe associations among observed composite scores rather than among validated latent constructs. Both associations were retained in the bootstrap, heteroskedasticity-consistent, and expanded models, but the magnitude of the usefulness coefficient depended on which other evaluative dimensions were included (Section 4.3). Self-rated knowledge of AI technologies, frequency of use, and comfort with AI use were not independently associated

with the outcome once the other variables were adjusted for. Because the knowledge dimension had low internal consistency, this pattern does not establish that experience matters more than knowledge. A small positive association with replacement concern was less stable across specifications and is treated as exploratory (Section 4.5). Exploratory group-specific analyses suggested that the association with experience was evident among students and that with replacement concern among practicing clinicians, although the status interactions were not jointly significant. After correction for multiple comparisons, students and clinicians differed on 1 of 11 dimensions. Because all measures were collected concurrently, these findings describe associations among appraisals held at one time point; they do not show that usefulness appraisals or experience lead to intended use, or that changing either would change intentions.

### *4.2. Experiential Engagement and Self-Rated Knowledge*

Experience with ChatGPT/AI was among the strongest bivariate correlates of likelihood of AI use ($\rho = 0.41$, $p < 0.001$), and its association persisted under multivariable adjustment ($\beta = 0.239$, $p < 0.001$), whereas those for self-rated knowledge, frequency of use, and comfort did not. Self-rated knowledge was modestly associated with the outcome at the bivariate level ($\rho = 0.21$, $p < 0.001$) but not once usefulness and experience were entered alongside it. Calls for AI preparation in the field have emphasized AI literacy [18,27], but the present data concern self-appraised familiarity rather than measured AI literacy, so no inference about objective knowledge, or about the value of knowledge-oriented preparation, can be drawn from them.

Three qualifications apply. First, the knowledge dimension indexed self-appraised familiarity rather than tested knowledge, and its internal consistency was low ($\alpha = 0.497$). Its null coefficient is therefore imprecise and does not indicate that AI knowledge is unrelated to intended use. Low reliability attenuates the bivariate association between a predictor and the outcome and, in a multivariable model, can bias its coefficient in either direction (Section 4.7). The 95% CI for the unstandardized coefficient $[-0.167, 0.148]$ was also compatible with small positive or negative associations, so the result indicates an absence of evidence for an independent association rather than evidence that none exists. Re-estimating the model with the more internally consistent two-item score did not change this result (Table S1). Reliability-corrected bivariate associations with likelihood of AI use (approximately 0.32 for knowledge and 0.46 for experience) and the dominance analysis (Table S8) ordered the two dimensions in the same way, but both depend on reliability estimates and score properties that a three-item dimension with $\alpha = 0.497$ supports poorly, and at the item level none of the knowledge items was independently associated with the outcome (Table S13); the ordering is therefore not interpreted as evidence that experience matters more than knowledge. Second, experience, but not frequency of use, retained an independent association. Both dimensions index use rather than evaluation: frequency captures general habitual use, whereas experience captures prior use across specified situations, including professional work. This pattern is consistent with situated use, rather than general frequency of use, being more closely related to intended professional use. Frequency scores were also concentrated near the top of the scale (71.4% of participants scored 4 or above), so restricted variance may have contributed to its null coefficient. Third, in exploratory group-specific models, the experience association was evident among students ($\beta = 0.409$) but weaker and imprecise among practicing clinicians ($\beta = 0.135$; Table S9). The difference between the two coefficients was itself not established: the experience × status interaction did not survive correction ($p = 0.033$; Bonferroni-corrected $\alpha = 0.0083$), and the six interaction terms did not jointly improve the model ($p = 0.186$). A coefficient that reaches significance in one group but not

in another does not show that the two coefficients differ, so this pattern is a hypothesis for independent samples and should not be generalized to both groups.

One possible account, offered as interpretation rather than as a tested claim, is practical calibration: using an AI tool for a concrete task may yield information about its capabilities and limits. One student stated this directly: "AI is a very useful and supportive element, but if the person has no professional knowledge it will be no different from a standard Google search engine" (ST4). The correlational data are compatible with this account, as experience correlated moderately with both comfort and frequency of use (both $\rho = 0.53$), but they cannot establish the order in which these develop. The UTAUT proposition that the influence of social factors on intention weakens with experience [10] is consistent with the same reading, although social influence was not measured here.

### *4.3. Perceived Usefulness and Technology Acceptance*

The association between perceived usefulness and likelihood of AI use ($\beta = 0.427$, $p < 0.001$) is the most theoretically anchored finding. It is qualified by the measurement structure of the usefulness score: a single factor fitted poorly unless residual dependence between two pairs of items was allowed, and even then RMSEA remained high and two items loaded weakly on the common factor (Section 3.2). The association was retained when usefulness and likelihood of use were modeled as latent variables (Table S12), but the composite combines heterogeneous ratings of usefulness, and its coefficient should not be read as an estimate for a validated construct. The association parallels the central relationship in TAM [9], although the usefulness score summarizes appraisals across areas of practice rather than measuring the TAM construct of perceived usefulness. It is also consistent with the usefulness–intention link being the most frequently confirmed relationship in healthcare technology acceptance research [11], and with meta-analytic evidence identifying performance expectancy as the strongest predictor of usage intention [2]. In medical education, Liu et al. [12] reported that attitudes mediated the associations of perceived usefulness, perceived ease of use, and perceived risk with students' intentions to use large language models for learning, and that perceived usefulness was not directly associated with intention once attitude was modeled. General attitude toward AI was not entered in the present model, so the usefulness coefficient may partly carry variance that an attitudinal construct would absorb. The present study also differs from that work in its outcome, which concerned intended use in professional practice rather than for learning, and in its inclusion of practicing clinicians. It is not a formal test of TAM or UTAUT: perceived ease of use, social influence, and facilitating conditions were not measured, and neither were trust or organizational support, which Lee et al. [1] argue require explicit modeling in clinical settings. Social influence may be particularly relevant: among US rehabilitation professionals, including SLPs, peer influence was the strongest predictor of generative AI use [14].

The magnitude of the usefulness association was specification-dependent. Entering four evaluative dimensions reduced the coefficient to $\beta = 0.276$ and identified perceived client satisfaction as an additional independent correlate (Table S2), so part of what the usefulness score captures is shared with beliefs about how clients would respond to AI. Because the measures were concurrent, it cannot be determined whether such beliefs sit alongside usefulness appraisals or between them and intended use. Usefulness and likelihood of use were also measured in the same questionnaire, so common-method variance may have strengthened their association. The two measures also share related content, in particular use during therapy sessions and intervention, and administrative work and report writing. In a two-factor model that fitted acceptably once residual dependence among usefulness items was allowed, the two dimensions were distinguishable (latent

correlation 0.65, 95% CI [0.55, 0.76]), but a residual association between the administrative-duties and report-writing items indicated overlapping content (Section 3.5). The usefulness association remained when the two overlapping usefulness items were removed ($\beta = 0.355$) and was present for every task, including those without a counterpart in the usefulness items (Table S14). Content overlap is therefore unlikely to account for the whole association, but it may have inflated its magnitude, and the evidence supports distinguishable but substantially related dimensions rather than fully distinct constructs.

The open-ended responses located the usefulness of AI primarily in preparatory and administrative work, such as material development, documentation, and workload reduction, rather than in the therapeutic encounter itself. This parallels the finding of Austin et al. [7] that AI use among SLP clinicians and students was largely restricted to administrative tasks, and the report by Şimşek et al. [27] that Turkish speech and language therapy students used ChatGPT more for academic purposes than for clinical reasoning or case analysis. Perceived usefulness for such tasks does not establish suitability for clinical decision-making, in which AI output requires verification against clinical knowledge and the characteristics of the individual client, and responsibility for the decision remains with the clinician; participants attached the same conditions to their endorsements (Section 3.6). Structured, safety-centred frameworks for clinician rating of large language model output have also begun to appear, such as the MELMA tool piloted in otolaryngology [39]; its applicability to speech and language therapy tasks has not been examined.

### *4.4. Students and Practicing Clinicians*

After correction, the groups differed only in perceived client satisfaction, which practicing SLTs rated more positively ($U = 13{,}382$, $p = 0.002$, $r_{rb} = 0.19$). This corresponds to a probability of superiority of approximately 0.60, that is, the probability that a randomly selected clinician rated client satisfaction higher than a randomly selected student, with tied ratings counted as one half (0.50 indicating no difference); the difference was statistically reliable but small, and its practical significance is limited. This may reflect informal observation of client responses; perceived client satisfaction is not a measured client outcome and may be subject to positive response bias. Students' higher replacement concern ($U = 13{,}786$, $p = 0.006$, $r_{rb} = -0.17$) and higher self-rated knowledge did not survive correction and are treated as descriptive. The greater concern among students is compatible with the prominence of job security among health science students' concerns [22], although that review did not include clinicians.

The absence of detectable differences on the remaining dimensions does not establish equivalence; the comparison was neither designed nor powered as an equivalence test. Two further features limit what the comparison can show about professional development. The practicing group was predominantly early-career (69.6% with <2 years of experience), so the contrast lies between late undergraduate training and early practice rather than between trainees and experienced clinicians; and a cross-sectional comparison of different individuals cannot show how perceptions change as students enter practice. Austin et al. [7] found no significant differences on any of the 11 scales across practicing SLPs, graduate students, and undergraduates. Whether AI-related appraisals change during the transition into practice or with longer clinical experience requires longitudinal designs that follow graduates into practice. A further constraint concerns measurement. Before the professional-impact items, respondents were instructed to answer with their current or future career as a speech and language therapist in mind, so students are likely to have referred to anticipated practice and clinicians to their current professional tasks, and identical items need not carry identical meaning in the two groups. Measurement invariance between the groups was not established: except for likelihood of AI use, the configural models fitted poorly,

including after respecification, so equality-constraint tests do not show that loadings or intercepts were equivalent (Section 3.2). Group differences, or their absence, may therefore partly reflect differences in item functioning rather than in the underlying appraisals. This applies also to perceived client satisfaction, the one dimension on which the groups differed after correction: the difference is a difference in observed scores, and whether it reflects a difference in the underlying appraisal remains uncertain.

### *4.5. Replacement Concern*

Replacement concern was positively associated with likelihood of AI use in the primary model ($\beta = 0.110$, $p = 0.010$), and in exploratory group-specific analyses this association was evident only among practicing clinicians ($\beta = 0.177$, $p = 0.002$; students, $\beta = 0.012$, $p = 0.849$). The association was small and the least stable in the model: in the expanded specification, its coefficient decreased, and its HC3-based $p$-value (0.048) lay close to the conventional threshold. It was also no longer significant when respondents who failed the attention check were included ($\beta = 0.076$, $p = 0.059$). Its direction is opposite to that in the source study, in which replacement concern correlated negatively with likelihood of using AI for SLP work ($\rho = -0.247$, $p = 0.012$) and with perceived usefulness ($\rho = -0.284$, $p = 0.004$), whereas the corresponding associations here were positive ($\rho = 0.19$, $p < 0.001$) or null ($\rho = 0.07$) [7]. Both studies used Spearman correlations, but three task labels shared by the likelihood-of-use and replacement-concern items were adapted, so measurement differences may contribute to the divergence. Outside the profession, concerns about the reliability of generative AI output were likewise not consistently associated with lower use intention among first-year university students [40].

Replacement concern correlated with perceived accuracy of AI-generated assessment scores ($\rho = 0.19$) and assessment results ($\rho = 0.14$), but not with perceived usefulness ($\rho = 0.07$) or perceived client satisfaction ($\rho = 0.03$). One possible account is that concern about replacement tracks appraisals of the technical capability of AI in assessment tasks, which could accompany both greater intended use and greater perceived threat. The decrease in the replacement-concern coefficient when the evaluative dimensions were added is compatible with, but does not demonstrate, partial overlap with these appraisals. The corresponding interaction term did not survive correction for multiple testing ($p = 0.044$; Bonferroni-corrected $\alpha = 0.0083$), so the group-specific results do not establish that the association differs between groups, but they suggest that the pooled coefficient reflects mainly the clinician subsample. The finding is best treated as a divergence to be examined in independent samples. It does not diminish the legitimacy of professional concern about AI-mediated task substitution, which has been identified as requiring serious attention [6,7,18].

### *4.6. Questions for Future Research*

The study did not evaluate any educational intervention or follow participants over time, so its findings do not show that any form of AI preparation is more effective than another, or how AI-related appraisals change as students enter practice. They do raise questions that can be tested. First, because perceived usefulness and experience with AI were associated with intended use whereas self-rated familiarity was not, studies using validated measures of AI knowledge and observed use could examine whether supervised, task-embedded engagement with AI, in which learners or clinicians appraise AI output against clinical knowledge and the needs of individual clients, changes the quality of AI-related judgments or intended and actual use compared with information-focused preparation. Second, participants' concerns about weakened clinical reasoning (CL3), generic output for individualized planning (CL2), and the effect of in-session use on the therapeutic relationship (ST1, CL1), together with reports that written institutional AI

policies were absent or unknown (Table S7), identify outcomes and conditions that such studies would need to address, alongside data privacy [6,7] and professional accountability. Third, longitudinal designs that follow students into practice are needed to determine whether AI-related appraisals change with clinical experience.

### *4.7. Strengths and Limitations*

The study assessed students and practicing clinicians with the same instrument, excluded inattentive responses using a criterion set before analysis, controlled for multiplicity in both comparison families, and examined the stability of the primary model with bootstrap, dominance, and group-specific sensitivity analyses.

Several limitations qualify the findings. The cross-sectional design precludes causal inference and does not establish temporal ordering among appraisals, experience, and intended use. Participants were recruited by convenience sampling through professional and university networks, as no accessible national sampling frame was available. The sample was predominantly female (90.0% of students; 95.4% of clinicians) and early-career (69.6% of clinicians with <2 years of experience), and clinicians were concentrated in special education and rehabilitation centers (68.8%), with hospital and university settings sparsely represented, so the findings may not generalize to more experienced clinicians or to other service settings. Hospital-based clinicians made up only 4.2% of the clinician subsample ($n = 10$). Adding years of experience and work setting to the clinician model left the coefficient pattern unchanged, but the small numbers of experienced and hospital-based clinicians limit what this check can show. The sample was obtained through department contacts, WhatsApp groups, and social media; the response rate cannot be estimated, and the distribution of respondents across departments and regions is unknown; the findings therefore describe this sample rather than the national population of students or clinicians.

All measures were self-reported and collected in one questionnaire, so social desirability and common-method variance may have inflated some associations, particularly between perceived usefulness and likelihood of use. These two dimensions share no items and differ in scope and response anchors, and a two-factor model distinguished them (latent correlation 0.65; Section 3.5), but two pairs of their content areas are close (see Section 2.3), so content overlap may have inflated the magnitude of their association. The outcome was self-reported intended use rather than observed or appropriate use. The administered Turkish questionnaire is not published with the article because the source instrument is copyrighted.

The Turkish adaptation followed established cross-cultural adaptation guidance but was not psychometrically validated as part of the original design. Post hoc single-factor models fitted poorly for most dimensions, including perceived usefulness, the dimension with the largest coefficient in the primary model, and post hoc respecification with content-based residual covariances improved but did not fully resolve this misfit. The dimension scores are therefore best read as composites of related ratings rather than as measures of established unidimensional constructs, and the regression coefficients describe associations among these composites. This is consistent with the design of several dimensions, whose items sample distinct areas of practice, but it also means that the scores depend on which areas are included and on their equal weighting. Because the configural models also fitted poorly for most dimensions, measurement invariance between students and clinicians was not established. Frequency of use was measured with two general items rather than five technology-specific ratings, three task labels shared by the likelihood-of-use and replacement-concern items were adapted, one perceived-usefulness item was expanded with examples, the client-satisfaction stem was rephrased for comprehensibility, and the two perceived-accuracy dimensions were harmonized to a common list of seven clinical

domains; these changes limit numerical comparison with the source study. The self-rated knowledge dimension had low internal consistency ($\alpha = 0.497$). Because all predictors contain measurement error, the direction of bias in any individual coefficient cannot be determined, and neither the null knowledge coefficient nor the usefulness coefficient should be treated as conservative. Differences in reliability across dimensions could also favor the more reliable dimensions when coefficient magnitudes are compared. The sensitivity analyses were conducted post hoc and used a different standard-error estimator from the primary model.

The single attention-check item cannot distinguish inattention from isolated error, and although excluded and retained respondents did not differ significantly in professional status, gender, or age group, these comparisons had limited power and students were numerically more frequent among excluded respondents; including the excluded respondents left the associations with perceived usefulness and experience unchanged but weakened that with replacement concern (Section 3.5). Adjusting for professional status left the pooled coefficient pattern essentially unchanged, but group-specific models diverged for experience and replacement concern; because the joint interaction test was not significant and each subsample was smaller than the combined sample, neither homogeneity nor heterogeneity across groups is established. Finally, the open-ended responses were typically brief and were coded by consensus review rather than independent duplicate coding; they contextualize rather than corroborate the quantitative findings.

## 5. Conclusions

Among speech-language therapy students and practicing clinicians in Türkiye, intended AI use in professional practice was most closely associated, in the primary model, with perceived usefulness of AI and direct experience with ChatGPT/AI, although the magnitude of the usefulness association depended on model specification and, in exploratory group-specific analyses, the experience association was evident among students but not clinicians. Because several dimension scores, including perceived usefulness, showed weak single-factor structure, these associations concern composite scores rather than validated constructs. Self-rated AI knowledge, frequency of use, and comfort were not independently associated with intended use once these variables were accounted for, although the knowledge dimension had low internal consistency. The results therefore do not indicate that experience matters more than knowledge. Students and clinicians did not differ detectably on most dimensions; this does not establish equivalence (measurement invariance between the groups was not established), and the cross-sectional comparison cannot show whether appraisals change as students enter practice. The small positive association between replacement concern and intended use was unexpected, evident only among clinicians in exploratory analyses, less stable in sensitivity analyses, and requires examination in independent samples. Because the design was cross-sectional, the findings do not show that usefulness appraisals or experience lead to intended use. Nor do they describe observed AI use or its clinical consequences, which were not assessed. The clinician subsample was predominantly early-career and based in special education and rehabilitation centers, so the findings should not be extended to more experienced or hospital-based clinicians without further study. Whether AI preparation that includes supervised, clinically grounded engagement with AI and explicit verification of its output affects AI-related judgment or use, in pre-registration education or continuing professional development, is a question for prospective and interventional studies.

**Supplementary Materials:** The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/healthcare14234015/s1, Table S1: Post hoc sensitivity regression using a two-item self-rated knowledge score; Table S2: Post hoc expanded sensitivity regression including four evaluative dimensions; Table S3: Summary of contextual adaptations to the source questionnaire; Table S4: Turkish source text for the open-ended responses quoted in the main text; Table S5: Spearman rank-order correlations among questionnaire dimensions ($N = 377$); Table S6: Mann–Whitney U comparisons between students and practicing SLTs ($N = 377$); Table S7: Summary of open-ended responses regarding the use of artificial intelligence and ChatGPT in speech and language therapy contexts; Table S8: Post hoc general dominance analysis of the primary regression model; Table S9: Post hoc group-specific regression models for likelihood of AI use; Table S10: Post hoc reliability and single-factor confirmatory factor analyses for each questionnaire dimension ($N = 377$); Table S11: Post hoc measurement invariance between students and practicing SLTs; Table S12: Post hoc sensitivity analyses of the primary regression model for likelihood of AI use; Table S13: Post hoc item-level analysis of the self-rated knowledge dimension ($N = 377$); Table S14: Post hoc task-level analysis of likelihood of AI use ($N = 377$); File S1: STROBE checklist for cross-sectional studies.

**Author Contributions:** Ş.S.Y. planned the study, collected the data, conducted the analyses, and drafted the manuscript. B.Ö. and A.Ö. contributed to study planning and to the verification of data collection and data analysis. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

**Institutional Review Board Statement:** The study was approved by the Social and Human Sciences Ethics Committee of Fenerbahçe University (decision no. 2026/8-4 FBU; 21 September 2026) before data collection, which took place between 22 September and 10 October 2026. The study was conducted in accordance with the Declaration of Helsinki.

**Informed Consent Statement:** All participants provided electronic informed consent before participation, which was voluntary and anonymous. The manuscript contains no identifiable individual data; open-ended responses are quoted anonymously.

**Data Availability Statement:** The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request. The wording of the source questionnaire items is not reproduced because it is subject to copyright; the adaptations are summarized in Table S3.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| AI | Artificial intelligence |
| CI | Confidence interval |
| HC3 | Heteroskedasticity-consistent standard errors, type 3 |
| OLS | Ordinary least squares |
| SLP | Speech-language pathology/pathologist |
| SLT | Speech-language therapist |
| STROBE | Strengthening the Reporting of Observational Studies in Epidemiology |
| TAM | Technology Acceptance Model |
| UTAUT | Unified Theory of Acceptance and Use of Technology |
| VIF | Variance inflation factor |

## References

1. Lee, A.T.; Ramasamy, R.K.; Subbarao, A. Understanding psychosocial barriers to healthcare technology adoption: A review of TAM and UTAUT frameworks. *Healthcare* **2025**, *13*, 250. [CrossRef] [PubMed]
2. Thanthrige, A.; Lu, B.; Sako, Z.; Wickramasinghe, N. Determinants of health care technology adoption using an integrated UTAUT and task technology fit model: Systematic review and meta-analysis. *J. Med. Internet Res.* **2025**, *27*, e64524. [CrossRef] [PubMed]

3. Tangadulrat, P.; Sono, S.; Tangtrakulwanich, B. Using ChatGPT for clinical practice and medical education: Cross-sectional survey of medical students' and physicians' perceptions. *JMIR Med. Educ.* **2023**, *9*, e50658. [CrossRef] [PubMed]
4. Alkhaaldi, S.M.I.; Kassab, C.H.; Dimassi, Z.; Oyoun Alsoud, L.; Al Fahim, M.; Al Hageh, C.; Ibrahim, H. Medical student experiences and perceptions of ChatGPT and artificial intelligence: Cross-sectional study. *JMIR Med. Educ.* **2023**, *9*, e51302. [CrossRef] [PubMed]
5. Zhang, R.; Zhao, Y.; Duan, L.; Pan, H. Artificial intelligence exposure, perceived needs, preferences, and concerns among endocrinologists and trainees in China: A convenience survey. *BMC Med. Educ.* **2026**. [CrossRef]
6. Aggarwal, K.; Ravi, R.; Yerraguntla, K. Artificial intelligence and patient care: Perspectives of audiologists and speech-language pathologists. *Intell. Based Med.* **2025**, *11*, 100214. [CrossRef]
7. Austin, J.; Benas, K.; Caicedo, S.; Imiolek, E.; Piekutowski, A.; Ghanim, I. Perceptions of artificial intelligence and ChatGPT by speech-language pathologists and students. *Am. J. Speech Lang. Pathol.* **2025**, *34*, 174–200. [CrossRef] [PubMed]
8. Cherrez-Ojeda, I.; Gallardo-Bastidas, J.C.; Robles-Velasco, K.; Osorio, M.F.; Velez Leon, E.M.; Leon Velastegui, M.; Pauletto, P.; Aguilar-Díaz, F.C.; Squassi, A.; González Eras, S.P.; et al. Understanding health care students' perceptions, beliefs, and attitudes toward AI-powered language models: Cross-sectional study. *JMIR Med. Educ.* **2024**, *10*, e51757. [CrossRef] [PubMed]
9. Davis, F.D. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Q.* **1989**, *13*, 319–340. [CrossRef] [PubMed]
10. Venkatesh, V.; Morris, M.G.; Davis, G.B.; Davis, F.D. User acceptance of information technology: Toward a unified view. *MIS Q.* **2003**, *27*, 425–478. [CrossRef]
11. AlQudah, A.A.; Al-Emran, M.; Shaalan, K. Technology acceptance in healthcare: A systematic review. *Appl. Sci.* **2021**, *11*, 10537. [CrossRef]
12. Liu, F.; Chang, X.; Zhu, Q.; Huang, Y.; Li, Y.; Wang, H. Assessing clinical medicine students' acceptance of large language model: Based on technology acceptance model. *BMC Med. Educ.* **2024**, *24*, 1251. [CrossRef] [PubMed]
13. Chew, K.S.; Wong, S.S.L.; Ooi, S.K.; Lim, Z.J.; Jensen, G.G.; Khairul Azri, N.N.Z.; Shanmuganathan, L. Sufficient and necessary conditions for ChatGPT adoption in medical education: A combined partial least square-structural equation modelling and necessary condition analysis. *BMC Med. Educ.* **2026**, *26*, 977. [CrossRef] [PubMed]
14. Clancy, M.J.; Pletcher, E.R.; Bell, A.; Pollen, T.; Dekerlegand, R. Acceptance and use of generative AI in clinical rehabilitation: An interdisciplinary survey of US rehabilitation professionals. *Arch. Rehabil. Res. Clin. Transl.* **2025**, *8*, 100571. [CrossRef] [PubMed]
15. Alotaibi, M.M.; Alotaibi, H.M. Acceptance and use of GenAI among medical and health sciences students in Saudi Arabia: An extended TAM study. *BMC Med. Educ.* **2026**, *26*, 1454. [CrossRef] [PubMed]
16. Brahmi, Z.; Mahyoob, M.; Al-Sarem, M.; Algaraady, J.; Bousselmi, K.; Alblwi, A. Exploring the role of machine learning in diagnosing and treating speech disorders: A systematic literature review. *Psychol. Res. Behav. Manag.* **2024**, *17*, 2205–2232. [CrossRef] [PubMed]
17. Cordella, C.; Marte, M.J.; Liu, H.; Kiran, S. An introduction to machine learning for speech-language pathologists: Concepts, terminology, and emerging applications. *Perspect. ASHA Spec. Interest Groups* **2025**, *10*, 432–450. [CrossRef] [PubMed]
18. Mytsyk, H.; Suchikova, Y. Stench of errors or the shine of potential: The challenge of (ir)responsible use of ChatGPT in speech-language pathology. *Int. J. Lang. Commun. Disord.* **2025**, *60*, e70088. [CrossRef] [PubMed]
19. Zheng, Q.; Choudhry, A.; Liu, Z.; Rabbani, P.; Hu, Y.; Olszewski, A.; Huang, Y.; Xiong, J. AI-enhanced speech-language intervention documentation: Opportunities and design goals. In *Artificial Intelligence in Education. AIED 2025*; Cristea, A.I., Walker, E., Lu, Y., Santos, O.C., Isotani, S., Eds.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2025; Volume 15882, pp. 132–140. [CrossRef]
20. Birol, N.Y.; Çiftci, H.B.; Yılmaz, A.; Çağlayan, A.; Alkan, F. Is there any room for ChatGPT AI bot in speech-language pathology? *Eur. Arch. Otorhinolaryngol.* **2025**, *282*, 3267–3280. [CrossRef] [PubMed]
21. Aggarwal, K.; Ravi, R.; Yerraguntla, K. Use of artificial intelligence tools by audiologists and speech-language therapists: An international survey of academicians. *J. Otol.* **2025**, *20*, 20–25. [CrossRef] [PubMed]
22. Derakhshanian, S.; Wood, L.; Arruzza, E. Perceptions and attitudes of health science students relating to artificial intelligence (AI): A scoping review. *Health Sci. Rep.* **2024**, *7*, e2289. [CrossRef] [PubMed]
23. Ehlert, H.; Lüdtke, U. The application of artificial intelligence in speech and language therapy: Attitudes and expectations. *Int. J. Lang. Commun. Disord.* **2026**, *61*, e70172. [CrossRef] [PubMed]
24. Lockwood, A.; Nisha, S.I.; Hendricks, A.E.; Meadan, H. Artificial intelligence adoption and related considerations: Survey of practicing school-based speech-language pathologists in Ohio. *Lang. Speech Hear. Serv. Sch.* **2026**, 1–13. [CrossRef] [PubMed]
25. Aggarwal, K.; Ravi, R.; Gunjawate, D.R.; Yerraguntla, K.; Kumar, K. Utilizing artificial intelligence in audiology and speech-language pathology education: Insights from ASLP students. *Work* **2026**, *85*, 895–901. [CrossRef] [PubMed]
26. Yaşa, İ.C.; Dölek, M.; Eyilikeder Tekin, S.; Kaya, A.S.; Akgün, P.; Yılmaz, S.D.; Tokalak, S. Artificial intelligence in speech and language therapy: A qualitative comparative analysis of clinical applications and outcomes. *Inquiry* **2026**, *63*, 00469580261445316. [CrossRef] [PubMed]

27. Simsek, A.; Kaya, T.; Yildirim, M.E. ChatGPT in undergraduate speech and language therapy education: Students' perspectives and experiences. *BMC Med. Educ.* **2026**, *26*, 481. [CrossRef] [PubMed]
28. von Elm, E.; Altman, D.G.; Egger, M.; Pocock, S.J.; Gøtzsche, P.C.; Vandenbroucke, J.P. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. *Lancet* **2007**, *370*, 1453–1457. [CrossRef] [PubMed]
29. Ölçme, Seçme ve Yerleştirme Merkezi (ÖSYM). *2023 Yükseköğretim Kurumları Sınavı (YKS) Yükseköğretim Programları ve Kontenjanları Kılavuzu*; ÖSYM: Ankara, Türkiye, 2023. Available online: https://www.osym.gov.tr/2023yuksekogretim-kurumlari-sinavi-yks-yuksekogretim-programlari-ve-kontenjanlari-kilavuzu (accessed on 30 October 2026).
30. Meade, A.W.; Craig, S.B. Identifying careless responses in survey data. *Psychol. Methods* **2012**, *17*, 437–455. [CrossRef] [PubMed]
31. Beaton, D.E.; Bombardier, C.; Guillemin, F.; Ferraz, M.B. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* **2000**, *25*, 3186–3191. [CrossRef] [PubMed]
32. The jamovi Project. *jamovi*, Version 2.7.38 [Computer Software]. 2026. Available online: https://www.jamovi.org (accessed on 16 November 2026).
33. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed.; Lawrence Erlbaum Associates: Hillsdale, NJ, USA, 1988.
34. Tabachnick, B.G.; Fidell, L.S. *Using Multivariate Statistics*, 7th ed.; Pearson: Boston, MA, USA, 2019.
35. Azen, R.; Budescu, D.V. The dominance analysis approach for comparing predictors in multiple regression. *Psychol. Methods* **2003**, *8*, 129–148. [CrossRef] [PubMed]
36. Rhemtulla, M.; Brosseau-Liard, P.É.; Savalei, V. When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychol. Methods* **2012**, *17*, 354–373. [CrossRef] [PubMed]
37. Putnick, D.L.; Bornstein, M.H. Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Dev. Rev.* **2016**, *41*, 71–90. [CrossRef] [PubMed]
38. Rosseel, Y. lavaan: An R package for structural equation modeling. *J. Stat. Softw.* **2012**, *48*, 1–36. [CrossRef]
39. Ahmadov, N.; Muradova, A.; Aliyeva, A. "MELMA" in otolaryngology: Medical evaluation of large language model answers. Clinician-rated scoring (MELMA-Q) and web-based auditing (MELMA-W) novel tools for AI assessment. *Eur. Arch. Otorhinolaryngol.* **2026**. [CrossRef] [PubMed]
40. Yu, C. Utility judgments, risk concerns, and generative AI use intention among first-year university students. *Front. Psychol.* **2026**, *17*, 1924148. [CrossRef] [PubMed]


**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.