An AI-Driven Personalized Treatment Pathway Engine for Multi-Symptom Digital Mental Health Platforms: Algorithmic Architecture, Privacy-Preserving Cross-Modal Synchronization, and Audit-Traceable Clinical Reasoning

다중 증상 디지털 정신건강 플랫폼을 위한 AI 기반 개인 맞춤형 치료 경로 엔진: 알고리즘 아키텍처, 프라이버시 보존 교차 모달 동기화, 그리고 감사 추적 가능한 임상 추론

Alex Kwon[1]

[1]Boston Neuromind LLC, Boston, MA, USA

Former Visiting Scholar, Harvard Graduate School of Education (Mind, Brain, and Education)

Correspondence: alex@bostonneuromind.com

Abstract

Background. Digital mental health interventions (DMHIs) have proliferated rapidly, yet a persistent gap exists between symptom assessment and personalized, session-level treatment planning. Most platforms surface diagnostic results without bridging to actionable, evidence-based treatment pathways that adapt to individual context. A second gap concerns the integration of conversational agents (bots) with clinical assessment modules: bot-collected data is clinically valuable but raises significant privacy concerns when shared without explicit user control. **Objective.** We present a six-algorithm engine that operationalizes the assessment-to-treatment pathway for a multi-symptom digital mental health platform (Boston Neuromind's NeuroCatchers + TalkCatcher dual-module architecture), spanning ten symptom domains (ADHD, depression, anxiety, sleep, burnout, PTSD, OCD, bipolar, autism, and peak performance). **Methods.** We specify (a) a unified relational database schema linking patient assessment, training library, recommendation reasoning, and session records; (b) Algorithm 1 (ALG-1), a six-component recommendation scorer integrating symptom match, biomarker match, evidence strength (Cochrane d-normalized), Fischer dynamic skill compatibility, practical fit, and personal history; (c) ALG-2, a Fischer-level-aware session plan generator that distributes skill progression across N sessions; (d) ALG-3, an adaptive replanner triggered by session count changes, goal modifications, or dropout signals; (e) ALG-4, a privacy-preserving Bot↔Catcher bridge with opt-in default-off semantics; (f) ALG-5, a clinician/client-facing custom training builder; and (g) ALG-6, a bilingual (Korean/English) result renderer producing clinical and patient versions in parallel. **Results.** The architecture provides end-to-end audit traceability: every recommendation persists score components, evidence citations (PMID/DOI), reasoning in both languages, and the model

version that generated it. ADHD serves as a Gold Template for subsequent symptom-domain replication. **Conclusion.** This design offers a generalizable, regulator-aligned (FDA SaMD, HIPAA, GDPR Article 9 and 22) framework for multi-symptom personalized digital mental health, demonstrating how recommendation algorithms can operate transparently while respecting privacy boundaries between conversational and clinical data streams.

Keywords: digital mental health; personalization; clinical decision support; adaptive treatment; privacy-preserving AI; explainable AI; Fischer dynamic skill theory; audit traceability

2. Introduction

2.1 The Digital Mental Health Landscape and the Personalization Gap

The past decade has produced a marked expansion of digital mental health interventions (DMHIs), including smartphone applications, web-based programs, conversational agents, and virtual reality experiences (Torous et al., 2021). Meta-analytic evidence indicates that app-supported interventions yield small-to-moderate effect sizes for depression, anxiety, and related conditions, with effect sizes consistently larger when interventions incorporate adaptive or personalized elements (Linardon et al., 2024). Personalization itself is a heterogeneous construct in the literature: Hornstein et al. (2023) identified at least six personalization dimensions in DMHIs for depression alone, while Wanniarachchi et al. (2025) catalogued 67 distinct personalization variables across 67 studies of DMHIs for youth depression and anxiety.

Yet a structural gap persists. The dominant pattern across deployed DMHIs is symptom screening followed by static psychoeducational content, generic skill exercises, or unstructured chatbot conversation. The translation from a diagnostic signal (e.g., a Patient Health Questionnaire-9 score, an ADHD self-report screener result, a biomarker pattern) to a personalized, ordered, evidence-anchored sequence of treatment activities — with explicit reasoning that can be audited by clinicians, patients, and regulators — remains largely unaddressed in production systems. This is the central problem the present work addresses.

2.2 From Assessment to Action: The Treatment Pathway Problem

A treatment pathway, as we use the term here, is an ordered, individualized sequence of evidence-based therapeutic activities that maps to a patient's symptom profile, biomarker signature (where available), developmental skill level, practical constraints, and personal history. Adaptive treatment strategy frameworks pioneered by Murphy (2005) and elaborated by Almirall et al. (2014) and Nahum-Shani et al. (2018) provide the conceptual scaffolding for sequencing decisions in mental health, but these frameworks do not specify the data structures or algorithmic mechanisms by which a real-world platform serving ten distinct symptom domains can operationalize the assessment-to-treatment translation at scale.

Clinical decision support systems (CDSSs) in physical medicine have demonstrated the value of algorithmic recommendation with persisted reasoning (e.g., Lee et al., 2024, for hepatocellular carcinoma; Sutton et al., 2020, for overview), but mental health CDSSs have lagged, in part because (a) symptom heterogeneity within a diagnostic category is large, (b) biomarker grounding is weaker than in oncology or cardiology, and (c) the patient's subjective preferences and developmental stage are themselves substantive inputs to the decision. We argue that these

obstacles, far from precluding algorithmic recommendation, instead define its required architecture: a multi-component scorer with explicit handling of subjective preference, developmental skill matching, and audit-traceable reasoning.

2.3 The Privacy Paradox in Bot-Clinical Integration

A second, less frequently articulated gap concerns the integration of conversational agents with clinical assessment modules. Generative AI chatbots have demonstrated clinical efficacy in the first wave of rigorous trials (Heinz et al., 2025; Wang et al., 2025; see also Reyes-Portillo et al., 2025, for a systematic review), and contemporary platforms increasingly deploy both a conversational front-end and a structured clinical back-end. The conversational stream contains information of considerable clinical value — mood trajectories, life events, adherence reports, symptom fluctuations — yet routing this stream into the clinical record without explicit user control creates significant ethical and regulatory exposure under the General Data Protection Regulation (GDPR) Article 9 (special category data) and Article 22 (automated individual decision-making), as well as the Health Insurance Portability and Accountability Act (HIPAA) in the United States (Cohen & Mello, 2018; Voigt & von dem Bussche, 2017; Wachter et al., 2017).

We frame this as the privacy paradox of bot-clinical integration: bot-derived clinical signal is valuable, yet unconsented sharing damages the trust premise on which conversational disclosure rests. The architectural solution we propose is an explicit opt-in default-off bridge between the conversational module (TalkCatcher) and the clinical assessment module (NeuroCatchers) — described in detail as ALG-4 below.

2.4 Audit Traceability as a Clinical and Regulatory Requirement

Recent regulatory trajectories converge on a shared requirement: AI-enabled clinical software must provide transparent, traceable reasoning for its recommendations. The U.S. Food and Drug Administration's AI/ML Software-as-a-Medical-Device (SaMD) action plan (FDA, 2021; Warraich et al., 2024) emphasizes good machine learning practice and lifecycle audit. The European Union's AI Act, which entered into force in 2024, classifies AI systems used in healthcare as high-risk and imposes transparency, human oversight, and post-market monitoring requirements. GDPR Article 22 affords data subjects a right not to be subject to solely automated decision-making with legal or similarly significant effects, accompanied by interpretive debate over the scope of an attendant "right to explanation" (Goodman & Flaxman, 2017; Wachter et al., 2017). Clinically, audit-traceable reasoning supports professional accountability under the standard of care doctrine, enables retrospective quality improvement, and equips clinicians to challenge or accept algorithmic recommendations on substantive grounds (Doshi-Velez & Kim, 2017; Ghassemi et al., 2021; Tjoa & Guan, 2021).

Accordingly, our architecture treats audit traceability as a first-class design constraint rather than a post-hoc add-on. Every recommendation produced by the engine persists its score components, evidence citations (with PubMed identifiers or digital object identifiers where available), reasoning text in both Korean and English, and the model version that generated it. We elaborate the audit schema in Section 12.

2.5 Contribution and Paper Organization

This paper makes the following contributions. First, we specify a six-algorithm engine — ALG-1 through ALG-6 — that collectively operationalizes the assessment-to-treatment pathway, the privacy-preserving bot-clinical bridge, custom training authoring, and bilingual result rendering. Second, we provide the relational database schema that underwrites these algorithms, with explicit attention to the audit trail. Third, we articulate a universal symptom-domain architecture: a single set of schemas and algorithms parameterized across ten symptom tracks (ADHD as the Gold Template, with depression, anxiety, sleep, burnout, post-traumatic stress disorder, obsessive-compulsive disorder, bipolar disorder, autism, and peak performance as parallel instantiations). Fourth, we map our design to current regulatory frameworks and identify the conformity pathway under FDA SaMD Class II and GDPR Article 9/22.

The remainder of the paper is organized as follows. Section 3 reviews related work in personalization, adaptive treatment, recommendation systems in healthcare, Fischer dynamic skill theory, privacy-preserving health data architecture, and audit trails in clinical AI. Section 4 describes the two-module system architecture. Section 5 presents the database schema. Sections 6 through 11 specify the six algorithms in turn. Section 12 elaborates the audit traceability layer. Section 13 describes the universal symptom architecture and the ADHD Gold Template strategy. Section 14 discusses strengths, limitations, clinical implications, and the regulatory pathway. Section 15 concludes.

3. Background and Related Work

3.1 Personalization in Digital Mental Health

Hornstein et al. (2023) provide the most influential recent framework for classifying personalization strategies in DMHIs, distinguishing dimensions of content, presentation, timing, and feedback. Wanniarachchi et al. (2025) extend this framework to youth populations and document the prevalence of each dimension. A consistent finding across these reviews is that few interventions implement just-in-time adaptive logic, and even fewer expose the algorithmic basis

of their personalization to users or clinicians. Our work addresses this gap by treating personalization as an explicit algorithmic process with persisted reasoning.

3.2 Conversational AI for Mental Health

Heinz et al. (2025) reported the first randomized controlled trial of a fine-tuned generative AI chatbot (Therabot) for clinical-level mental health symptoms, with N=210 adults across major depressive disorder, generalized anxiety disorder, and clinically high-risk feeding and eating disorders. Effect sizes were large and engagement was high (mean 260 messages per user over 4 weeks). Wang et al. (2025) extended the evidence base to cognitive behavioral therapy chatbots for depression in Chinese university students. Reyes-Portillo et al. (2025) provide a meta-analytic synthesis. These trials establish the clinical viability of conversational AI but do not specify how conversational content should be safely bridged to a structured clinical record — a gap our ALG-4 addresses directly.

3.3 Recommendation Systems in Healthcare

Lee et al. (2024) provide a recent npj Digital Medicine exemplar of a machine-learning CDSS that produces treatment recommendations with ranked alternatives and survival predictions for hepatocellular carcinoma. Sutton et al. (2020) review the broader CDSS landscape, identifying benefits (decision support, reduced errors) and risks (alert fatigue, automation bias). Wiens et al. (2019) articulate a "do no harm" roadmap for responsible ML in clinical contexts. Our recommendation engine (ALG-1) inherits the multi-component scoring tradition while extending it with explicit developmental and preference components calibrated for mental health.

3.4 Adaptive Treatment Strategies

Murphy's (2005) SMART (Sequential Multiple Assignment Randomized Trial) framework formalizes the multi-stage decision problem of adaptive treatment. Almirall et al. (2014) extend SMART to behavioral interventions including weight loss; Nahum-Shani et al. (2018) elaborate just-in-time adaptive interventions for mobile health. While SMART trials specify the empirical methodology for discovering adaptive strategies, they do not prescribe the data architecture for operationalizing such strategies in a production platform. ALG-2 and ALG-3 together provide one such operationalization, integrating session-level planning with re-planning triggered by patient-side changes.

3.5 Fischer's Dynamic Skill Theory

Fischer's (1980) dynamic skill theory, elaborated in Fischer and Bidell (2006) and Mascolo and Fischer (2015), posits that skills develop in a hierarchical sequence across four tiers (reflexes,

sensorimotor actions, representations, abstractions), with four levels per tier yielding a developmental yardstick. Skills are conceptualized as contextualized control structures rather than general abilities, and skill performance varies as a function of contextual support. We operationalize Fischer's framework in ALG-1 as a fischer_compatibility component (matching training entry level to estimated current level) and in ALG-2 as a session-by-session progression schedule. This is, to our knowledge, the first explicit computational use of Fischer's framework in a deployed mental health platform.

3.6 Privacy-Preserving Health Data Architecture

Voigt and von dem Bussche (2017) provide the standard practical guide to GDPR compliance, with particular emphasis on Article 9 (special category data, including health) and Article 22 (automated decision-making). Cohen and Mello (2018) examine HIPAA modernization in light of contemporary data flows. The privacy paradox we articulate in Section 2.3 maps to GDPR Article 9(2)(a)'s requirement for explicit consent prior to processing special category data, which is the legal basis for the opt-in default-off architecture in ALG-4.

3.7 Explainability and Audit in Clinical AI

Doshi-Velez and Kim (2017) articulate the scientific foundations of interpretable ML; Tjoa and Guan (2021) survey explainable AI methods with explicit medical applications; Ghassemi et al. (2021) caution against over-reliance on post-hoc explanation in healthcare, arguing instead for transparent design from the outset. Wachter et al. (2017) examine the contested scope of a GDPR-derived "right to explanation," while Goodman and Flaxman (2017) offer a more expansive reading. Our audit traceability layer (Section 12) embraces the transparent-by-design stance, persisting the substantive inputs to each recommendation rather than relying on post-hoc explanation methods.

4. System Overview

Figure 1. Two-Module Architecture with Six-Algorithm Integration

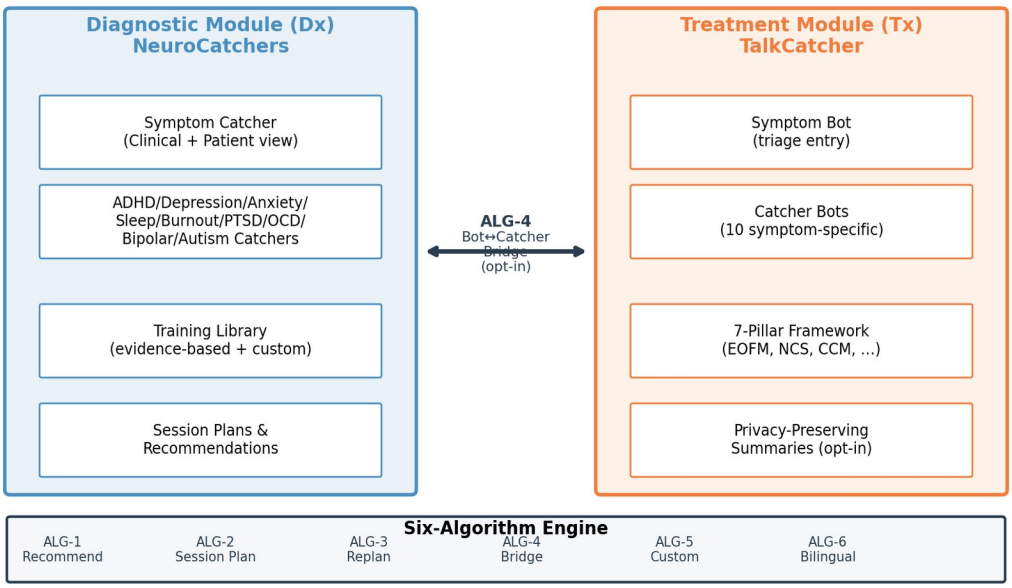

Figure 1. Two-module system architecture: NeuroCatchers (Diagnostic) and TalkCatcher (Treatment) linked by ALG-4 privacy-preserving bridge.

4.1 Two-Module Architecture: Diagnostic and Treatment

The platform is organized as two complementary modules (Figure 1). The Diagnostic Module (Dx), branded NeuroCatchers, provides structured clinical assessment via a top-level Symptom Catcher and ten symptom-specific Catchers (ADHD, depression, anxiety, sleep, burnout, post-traumatic stress disorder, obsessive-compulsive disorder, bipolar disorder, autism, and peak performance). Each Catcher contains an evidence-based training library and surfaces session plans and recommendations. The Treatment Module (Tx), branded TalkCatcher, provides conversational support via a Symptom Bot (triage entry) and ten Catcher Bots corresponding to the clinical symptom domains. The conversational module instantiates Boston Neuromind's seven-pillar framework, which includes the Emotion-Only Feedback Model (EOFM), the Narrative Co-Construction System (NCS), the Clinical Calibration Module (CCM), and four additional pillars not central to the present paper.

The two modules are connected by ALG-4, a privacy-preserving bridge that synchronizes bot-derived clinical signal into the diagnostic record only when the patient has explicitly opted in. This separation reflects the privacy paradox articulated in Section 2.3.

4.2 Six-Algorithm Engine

Six algorithms collectively operationalize the assessment-to-treatment pathway and the bridge between modules. ALG-1 (Recommendation Engine) takes patient profile and candidate trainings as input and produces ranked recommendations with reasoning. ALG-2 (Session Plan Generator) takes a selected training and produces an N-session plan with Fischer-level progression. ALG-3 (Adaptive Replan) updates the plan in response to triggering events. ALG-4 (Bot↔Catcher Bridge) governs privacy-preserving synchronization. ALG-5 (Custom Training Builder) lets clinicians or patients author novel training records that conform to the universal schema. ALG-6 (Bilingual Result Renderer) produces Korean and English versions of diagnostic outputs.

4.3 Universal Symptom-Domain Structure

A core architectural commitment is universality across symptom domains. A single set of database tables, a single set of algorithms, and a single rendering pipeline serve all ten Catchers. Symptom-specific content is data, not code: the training_library table contains a symptom_track column that selects which trainings are visible for a given Catcher; the contraindications field encodes symptom-specific safety rules; the quick_input_schema specifies symptom-appropriate measurement instruments. We adopt ADHD as the Gold Template for initial development and validation, with the explicit intent that nine subsequent symptom domains be instantiated by populating data rather than writing code.

4.4 Privacy-First Design Principles

Five design principles govern the architecture. First, opt-in default-off semantics apply to all cross-module data flow. Second, consent is granular and withdrawable, consistent with GDPR Article 7. Third, audit traceability is a first-class output of every recommendation. Fourth, bilingual rendering is mandatory for diagnostic outputs to serve both Korean and English-speaking populations. Fifth, the architecture must accommodate both autonomous patient use and clinician-supervised use, with the same data structures supporting both modes.

5. Database Schema Design

Figure 2. Core Database Schema — Entity Relationship Diagram

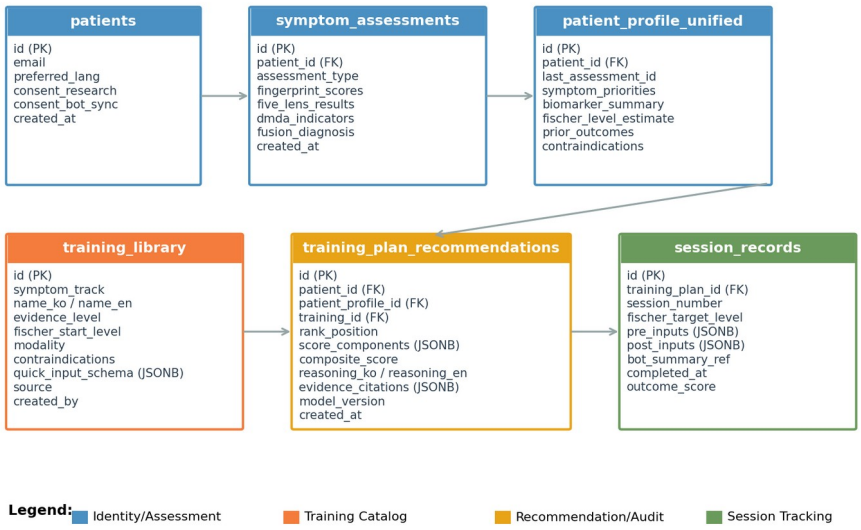

Figure 2. Entity-relationship diagram showing six core tables and their relationships.

Table 1. Core database tables and relationships.

Table	Primary Purpose	Key Fields	Used By
patients	Identity and consent flags	id, language_pref, consent_research, consent_bot_sync	All algorithms
symptom_assessments	Raw assessment outputs	id, patient_id, fingerprint_jsonb, dmda_jsonb	ALG-6
patient_profile_unified	Synthesized profile for recommendation	symptom_priorities, biomarker_summary, fischer_level_estimate	ALG-1
training_library	Curated intervention catalog	id, symptom_track, evidence_level, fischer_start_level, contraindications, quick_input_schema	ALG-1, ALG-2, ALG-5
training_plan_recommendations	Persisted recommendation audit	score_components, composite_score, reasoning_ko, reasoning_en, evidence_citations, model_version	ALG-1, audit layer

Table	Primary Purpose	Key Fields	Used By
session_records	Per-session tracking	pre_input, post_input, bot_summary, completion_status	ALG-2, ALG-3, ALG-4

5.1 Core Entities and Relationships

The schema comprises six core tables (Figure 2). The patients table holds identity, language preference, and consent flags including consent_research and consent_bot_sync. The symptom_assessments table stores results from clinical instruments, organized as JSONB columns for the Fingerprint Engine output, Five-Lens scores, Differential Mental Disorder Algorithm (DMDA) indicators, and the fused diagnostic profile. The patient_profile_unified table consolidates the most recent assessment with a synthesized symptom priority vector, biomarker summary, an estimated Fischer skill level, and contraindication flags; this table is the primary input to ALG-1.

The training_library table catalogs evidence-based interventions (Section 6). The training_plan_recommendations table persists every ALG-1 output with full score decomposition and reasoning. The session_records table tracks per-session activity, including pre-session and post-session quick inputs and optional bot summary references.

5.2 The patient_profile_unified Table as Recommendation Input

A central design choice is the consolidation of heterogeneous assessment outputs into a single unified profile that serves as the input to recommendation. The unified profile contains: symptom_priorities (a normalized vector indicating relative weighting across symptom domains), biomarker_summary (where electroencephalographic or other physiological data is available), fischer_level_estimate (the patient's estimated current developmental level), prior_outcomes (historical engagement, dropout, and outcome data), and contraindications (a set of flags barring certain interventions). This consolidation enables ALG-1 to operate against a stable, well-defined input rather than navigating raw assessment data.

5.3 The training_library Table: Evidence Tier and Schema

Each row in training_library represents one therapeutic intervention. Critical columns include: id, symptom_track (which Catcher surfaces this training), name_ko and name_en, evidence_level (one of rct, meta_analysis, consensus_guideline, expert_opinion, custom), fischer_start_level (the developmental entry point), modality (cognitive, behavioral, somatic, neurofeedback, exposure, mindfulness, psychoeducation), contraindications (JSONB), quick_input_schema (JSONB specifying per-session measurement instruments), and source (which clinician or

organization curated the entry). The dual evidence_level + custom design allows curated, vetted entries to coexist with clinician-authored bespoke trainings without compromising the recommendation engine's evidence-strength scoring.

5.4 Recommendation Reasoning Persistence

Every ALG-1 invocation writes one or more rows to training_plan_recommendations. Critical fields include score_components (JSONB containing all six component scores with their sub-components and methods), composite_score (the final weighted total), reasoning_ko and reasoning_en (human-readable explanations in both languages), evidence_citations (JSONB array of PubMed identifiers or digital object identifiers with the specific claims they support), contraindications_considered (which patient flags were evaluated and how), and model_version (the identifier of the scoring model that produced this recommendation). The model_version field is essential for longitudinal audit: as the engine learns and is updated, all historical recommendations remain interpretable in the context of the model that generated them.

5.5 Bot-Catcher Synchronization Schema

When consent_bot_sync is TRUE, ALG-4 writes per-session summaries into session_records.bot_summary, a JSONB field with a fixed schema: mood_trend (rolling average over a configurable window), reported_symptoms (an array of clinically relevant signals extracted from conversation), life_events (significant events the user disclosed, such as moves or job changes), and adherence_self_report (the user's stated engagement with prescribed activities). The schema is intentionally narrow to support clinical decision-making while minimizing unnecessary data flow.

6. Algorithm 1: Recommendation Engine

6.1 Input and Hard Filters

ALG-1 accepts as input a row from patient_profile_unified and the subset of training_library entries whose symptom_track matches the active Catcher. Before scoring, hard filters apply: any training whose contraindications array intersects the patient's contraindication flags is removed from the candidate set. Hard filters are not weighted; they categorically exclude trainings that are clinically inappropriate for the specific patient. Examples include medication interactions (a stimulant-based protocol is filtered when the patient flags a contraindicated cardiac condition), comorbidity exclusions (an exposure protocol for OCD is filtered when active psychosis is flagged), and developmental floor (a high-Fischer-level training is filtered when the patient's estimated level falls below the training's fischer_start_level minus an admissible gap).

6.2 Six-Component Scoring Function

Surviving candidates are scored across six components (Figure 3, Table 2). The symptom_match component computes a weighted cosine similarity between the patient's symptom_priorities vector and the training's symptom_target_vector. The biomarker_match component, applicable when biomarker_summary is populated, computes a Z-score-weighted alignment between patient biomarker signature and training's biomarker_target_pattern. The evidence_strength component normalizes Cochrane d (or analogous effect size) from the training's source evidence; meta-analytic effect sizes receive higher weight than single-trial estimates, and consensus-guideline trainings receive a calibrated baseline.

The fischer_compatibility component scores the match between the patient's fischer_level_estimate and the training's fischer_start_level, with maximum score at exact match and graceful degradation for gaps. The practical_fit component aggregates modality preferences, time-availability constraints, and cost factors. The personal_history component rewards trainings related to prior successes and penalizes trainings related to prior dropouts. Each component yields a score in the unit interval [0, 1] with a documented derivation method.

6.3 Composite Scoring and Ranking

The composite_score is computed as a weighted sum of the six component scores. Default weights are derived from clinician consultation and may be adjusted per symptom track. Candidates are ranked by descending composite_score, and the top three are presented as primary recommendations; the fourth through tenth are retained as alternates and surfaced upon user request. The composite weights themselves are stored in the model_version metadata to support audit.

6.4 Reasoning Generation

For each ranked candidate, ALG-1 generates Korean and English reasoning text via a template-plus-large-language-model-polish pipeline. Templates are written by clinicians and include placeholders for the patient's salient features (e.g., "Given your reported priority of attention regulation and your estimated developmental level of sm3..."). The LLM polish step ensures tonal coherence and natural phrasing while preserving the factual content from the template. Generated reasoning is constrained never to fabricate evidence citations; only PMIDs and DOIs that were present in the source training's evidence_citations field may appear in the rendered reasoning.

6.5 Cold-Start Handling

When a patient has no prior history (cold-start), the personal_history component defaults to a neutral 0.5 and the composite weight on this component is temporarily redistributed across the

remaining five components. Similarly, when biomarker_summary is empty, biomarker_match is omitted and its weight is redistributed. The redistributions are recorded in score_components so that comparative analyses across cohorts can correctly account for the data available at each recommendation.

6.6 Example: ADHD Recommendation Output

Table 3 illustrates a sample output for an adult ADHD patient with mild inattention symptoms, no biomarker data, and an estimated Fischer level of sm3. The first-ranked recommendation is a focused-attention neurofeedback protocol with composite_score 0.89, supported by Arns et al. (2009) and Van Doren et al. (2019). The second is a Pomodoro-based behavioral protocol (composite_score 0.84). The third is a mindfulness-based attention training (composite_score 0.81). Each is accompanied by Korean and English reasoning text and the relevant evidence citations.

7. Algorithm 2: Session Plan Generator

7.1 Input and Fischer-Level Progression

ALG-2 accepts a selected training_library row, an N_sessions count chosen by the patient or clinician, and the patient's current fischer_level_estimate. It produces an ordered sequence of N session specifications, with Fischer level progressing from the training's start level to an end level appropriate to the patient's developmental trajectory and the training's structure (Figure 4, Table 4).

The progression algorithm respects Fischer's tier boundaries: transitions within a tier (sensorimotor, representational, abstract) occur more rapidly than tier-crossing transitions, which require additional consolidation sessions. The algorithm draws on Fischer and Bidell's (2006) characterization of skill construction as recurrent rather than strictly linear: practice at a given level consolidates and stabilizes the skill before progression.

7.2 Per-Session Content Generation

For each session, ALG-2 generates a content specification including: target Fischer level, target skill, activity description, expected duration, and the quick_input_schema instantiation that the patient will complete pre- and post-session. The quick_input_schema is inherited from the training_library row but parameterized to the session's specific focus. For example, an ADHD focused-attention training might use a 0-10 slider for pre-session focus rating, an integer count for tasks completed during the session, and a 0-10 slider for post-session focus rating; these labels are localized to Korean or English based on patient preference.

7.3 30-Second Session Entry Design

A key design constraint is that per-session data entry must be completable within thirty seconds. Adherence to digital interventions degrades rapidly when input burden exceeds this threshold (Baumel et al., 2019; Fleming et al., 2018). The quick_input_schema is consequently restricted to a maximum of four fields per session, each of which uses one of four primitive input types (slider, integer, boolean, time). Free-text entry is optional and never required for plan continuation.

8. Algorithm 3: Adaptive Replan

8.1 Trigger Conditions

ALG-3 fires in response to three classes of triggers (Figure 5). First, an explicit user or clinician action that changes the session count, modifies the primary goal, or selects a different training from the alternates ranked by ALG-1. Second, a dropout signal: missed sessions beyond a configurable threshold (default: two consecutive misses), or a deterioration in pre/post outcome scores beyond a configurable tolerance. Third, a periodic review at a configurable cadence (default: every six sessions) that re-evaluates the plan against the patient's updated profile.

8.2 Plan Diff and History Preservation

When ALG-3 fires, it does not modify the existing plan in place; instead, it records a plan_modification row with a diff between the prior plan and the new plan, then writes a new plan. This preserves the complete history of the patient's treatment trajectory, supporting both clinical reflection and longitudinal analysis. Historical plans remain queryable indefinitely.

8.3 Reasoning Regeneration

Each replan invocation regenerates the bilingual reasoning text via the same template + LLM polish pipeline used by ALG-1, with the addition of an explicit "what changed and why" narrative. The narrative cites the trigger (e.g., "extended from 10 to 12 sessions at the patient's request" or "alternative training selected due to two consecutive missed sessions of the prior plan") and the relevant updated profile components.

9. Algorithm 4: Privacy-Preserving Bot↔Catcher Bridge

9.1 The Privacy Paradox in Integrated Platforms

As articulated in Section 2.3, the integration of conversational agents with clinical assessment modules confronts a privacy paradox: bot-derived clinical signal is valuable, yet unconsented sharing damages the trust premise on which conversational disclosure rests. The choice between

full integration (high clinical utility, low privacy protection) and full separation (high privacy protection, low clinical utility) is a false dichotomy. ALG-4 implements a third option: explicit, granular, withdrawable opt-in with a default-off baseline.

9.2 Opt-In Architecture

The patients.consent_bot_sync flag governs synchronization (Figure 6). Its default value at account creation is FALSE. The flag may be toggled by the patient at any time through the privacy settings interface; toggling to FALSE immediately halts further synchronization but does not retroactively delete previously synchronized summaries (which are subject to standard data deletion requests under GDPR Article 17 and HIPAA right-of-access provisions). When a clinician is involved in the patient's care, the clinician may request that the patient enable sync, but the patient retains sole authority over the flag's value, in conformity with GDPR Article 9(2)(a)'s explicit consent requirement for special category data.

9.3 Summary Extraction Pipeline

When sync is enabled, the bot's conversation history is processed at session end (or, for batch operation, daily) through a retrieval-augmented generation pipeline anchored to a curated clinical vocabulary. The pipeline extracts content matching the bot_summary schema (Section 5.5) and discards content outside this schema. Critically, the raw conversation log is never written to the diagnostic record; only the schema-conformant extraction is transferred. This minimizes data flow consistent with GDPR Article 5(1)(c) data minimization.

9.4 Therapist Visibility Controls

Within the diagnostic record, bot_summary fields are subject to standard therapist access controls. Patients may further restrict specific bot_summary fields from therapist view (e.g., concealing life_events while permitting mood_trend); these field-level restrictions are recorded in an access_policy JSONB column on the session_records row. All accesses to bot_summary are logged for audit.

10. Algorithm 5: Custom Training Builder

10.1 Clinician and Client Input UI

Clinicians frequently develop bespoke protocols for individual patients (e.g., a graded exposure plan for a specific OCD ritual; a personalized morning routine for executive function support). ALG-5 supports authoring such trainings through an eight-field UI (Figure 7): symptom track, target behavior, category, two measurement specifications (each selecting from a constrained

dropdown), recommended sessions, per-session minutes, and visibility (clinician-only versus private). Free-form JSON authoring is explicitly disallowed; the constrained UI ensures that custom trainings conform to the universal schema and remain compatible with ALG-1 scoring.

10.2 AI-Assisted Refinement

An optional "Refine with AI" action invokes a language model to: generate a clinically appropriate description, propose contraindications drawn from the symptom track's standard contraindication library, and surface similar existing trainings for the author's reference. The author retains full editorial control over the AI-generated content. All AI-generated text is flagged in the source field of the training_library row.

10.3 Validation and Storage

Before insertion, validation checks: (a) all required fields are populated, (b) the contraindications do not conflict with the symptom track's mandatory inclusions, (c) the quick_input_schema yields valid input types. Custom trainings are stored with evidence_level = 'custom' and source recording the authoring clinician or client. They are eligible for ALG-1 scoring with a calibrated evidence_strength baseline below that of meta-analytically supported trainings.

11. Algorithm 6: Bilingual Result Renderer

11.1 Two Output Paths for Diagnostic Results

Symptom Catcher outputs are rendered in two parallel versions: a clinical version employing DSM/ICD codes, Z-scores, and neurological interpretation, and a patient version employing everyday language, metaphor, and encouragement (Figure 8). Symptom Bot outputs, by contrast, are rendered in a single unified version: the bot's communication context makes the clinical/patient distinction less meaningful, and the bot's tone calibration is already constrained by ALG-6's parameters.

11.2 Template Plus LLM Polish Pipeline

Both paths use a template + LLM polish pipeline analogous to ALG-1's reasoning generation. Templates encode the structural elements (which findings are reported, in what order, with what emphasis); the LLM polish step calibrates tone and ensures natural phrasing in each language. Tone calibration parameters vary by output: clinical version uses precise, professional register; patient version uses warm, encouraging register; bot version balances accessibility with diagnostic substance.

11.3 Korean-English Parallel Rendering

Both languages are rendered in parallel rather than via post-hoc translation. This avoids two failure modes of post-hoc translation: idiom preservation (Korean idioms do not directly translate to English equivalents, and vice versa) and clinical-term equivalence (Korean clinical terminology occasionally lacks direct English analogs and requires explanatory rendering). The template engine maintains separate templates for each language, sharing only the data inputs.

11.4 Cultural Calibration

Per Hwang et al. (2008), cultural context shapes both presentation of symptoms and acceptability of treatment recommendations. The Korean rendering employs honorifics appropriate to clinical encounters, somatic framings of psychological distress where culturally indicated, and family-system framings where individual-frame language might be culturally dissonant. These adaptations are encoded in the templates, not introduced post-hoc.

12. Audit Traceability

Table 6. Audit field schema in training_plan_recommendations.

Field	Type	Purpose
score_components	JSONB	All 6 component scores + sub-components + methods
composite_score	Numeric (0-1)	Final weighted total
reasoning_ko	Text (immutable)	Korean human-readable explanation
reasoning_en	Text (immutable)	English human-readable explanation
evidence_citations	JSONB array	PMID/DOI with claim and component link
contraindications_considered	JSONB	Which patient flags evaluated and outcome
model_version	Text	Identifier of scoring model that produced recommendation
created_at	Timestamp	Creation time
patient_id	UUID	Foreign key to patients

Table 7. Regulatory mapping of audit design.

Regulation	Requirement	How Audit Design Addresses It
FDA SaMD Good Machine Learning Practice	Transparency, lifecycle audit	score_components + model_version persistence
FDA SaMD Post-Market Monitoring	Tracking real-world performance	training_plan_recommendations + session_records linked
GDPR Article 9 (Special Category Data)	Explicit consent for health data	consent_bot_sync flag; opt-in default-off in ALG-4
GDPR Article 22 (Automated Decision-Making)	Right not to be subject to solely automated decisions	Clinician-in-the-loop design; reasoning_ko/en for meaningful review
GDPR Article 5(1)(c) Data Minimization	Limit data flow to necessary	Fixed bot_summary schema; raw conversation never transferred
EU AI Act High-Risk System	Transparency, human oversight, post-market monitoring	Audit fields + clinician collaboration model
HIPAA Privacy Rule	Patient control over PHI	Granular consent flags; withdrawal supported

12.1 First-Class Audit Design

As articulated in Section 2.4, we treat audit traceability as a first-class design constraint. The training_plan_recommendations table persists, for every recommendation, all six score components and their sub-components, the composite score, reasoning text in both Korean and English, evidence citations with PubMed identifiers or digital object identifiers where available, the contraindications considered (and the conclusion reached for each), and the version identifier of the scoring model that produced the recommendation. This design exceeds the minimum requirements of the U.S. Food and Drug Administration's Artificial Intelligence / Machine Learning Software-as-a-Medical-Device action plan (FDA, 2021) and aligns with the lifecycle audit emphasis in Warraich et al. (2024).

12.2 Score Component Persistence

The score_components JSONB column stores not only the numeric value of each component but also the sub-components and the method used to compute the value. For example, the symptom_match component records the cosine similarity value, the weight vector applied, the symptom_priorities slice extracted from the patient profile, and the matching slice from the training catalog. This level of detail enables retrospective replication: given the persisted record and the model_version, an auditor can recompute the score without access to the live engine state.

12.3 Evidence Citation Schema

The evidence_citations field is a JSONB array. Each entry contains a pmid or doi, a brief description of the claim the citation supports, and the relationship of that claim to the recommendation. For example, an ADHD neurofeedback recommendation might cite Van Doren et al. (2019) for sustained-effect evidence, with the claim "neurofeedback for ADHD shows medium effect size at six-month follow-up" linked to the evidence_strength component. This schema supports both clinician verification at the point of care and external audit during regulatory review.

12.4 Bilingual Reasoning Immutability

The reasoning_ko and reasoning_en fields are written at recommendation time and are immutable thereafter. Subsequent recommendations for the same patient generate new rows rather than updating existing ones. This design choice preserves the historical record needed for longitudinal audit and for understanding how the engine's reasoning has evolved over a patient's course of treatment. Bilingual immutability is essential for serving Korean and English-speaking populations under a single audit framework.

12.5 Regulatory Mapping

Table 7 maps the audit fields to specific regulatory requirements. The U.S. Food and Drug Administration's Software-as-a-Medical-Device guidance, Good Machine Learning Practice principles, the General Data Protection Regulation Article 22 right not to be subject to solely automated decision-making, and the European Union AI Act high-risk system transparency requirements are all addressed by the persistence design. We argue that this design is necessary but not sufficient for full conformity; outcome validation across diverse populations and post-market monitoring remain required prior to any clearance pursuit.

13. Universal Symptom Architecture

Table 8. Ten symptom domains and primary training categories.

Symptom Domain	Primary Training Categories	Key Evidence Source
ADHD (Gold Template)	Neurofeedback, CBT, mindfulness, environmental modification, skills training	Cortese et al. (2018); Van Doren et al. (2019)
Depression	CBT, behavioral activation, interpersonal therapy, mindfulness-based cognitive therapy	Cuijpers et al. (2020)
Anxiety	Exposure therapy, CBT,	Carpenter et al. (2018)

Symptom Domain	Primary Training Categories	Key Evidence Source
	mindfulness, relaxation training	
Sleep	CBT-I, sleep hygiene, stimulus control, sleep restriction	Riemann et al. (2023)
Burnout	Workload management, recovery activities, boundary-setting, values clarification	Maslach & Leiter (2016)
PTSD	Trauma-focused CBT, prolonged exposure, EMDR, psychoeducation	Bisson et al. (2020)
OCD	Exposure and response prevention, CBT, acceptance-based approaches	Skapinakis et al. (2016)
Bipolar	Psychoeducation, mood monitoring, social rhythm therapy, family-focused therapy	Goodwin et al. (2016)
Autism	Social skills training, sensory regulation, communication support, family support	Lord et al. (2022)
Peak Performance	Goal-setting, attention training, performance routine development	Fischer & Bidell (2006); skill-based literature

13.1 ADHD as Gold Template

We have chosen attention-deficit/hyperactivity disorder (ADHD) as the Gold Template for initial development and clinical validation. The choice reflects three considerations. First, ADHD has the strongest convergence between subjective symptom report, neuropsychological assessment (continuous performance tasks, n-back tasks), and biomarker data (quantitative electroencephalography theta/beta ratio; Arns et al., 2009; Van Doren et al., 2019), which allows full exercise of the six-component scoring scheme. Second, evidence-based interventions for ADHD span pharmacotherapy, cognitive-behavioral therapy, neurofeedback, mindfulness, environmental modification, and skills training (Cortese et al., 2018), allowing the recommendation engine to navigate genuine variety. Third, ADHD captures both pediatric and adult populations and presents across a Fischer skill range from early concrete operations to advanced abstractions, exercising the developmental-matching component.

The Gold Template strategy commits to deep clinical validation of ADHD before replication to other symptom domains. Validation targets include recommendation alignment with expert clinician choice (target: kappa ≥ 0.70), reasoning quality as judged by external clinical reviewers

(Likert scale, target mean ≥ 4 of 5), and patient comprehensibility of bilingual reasoning text (target mean ≥ 4 of 5 across Korean and English samples).

13.2 Symptom-Domain Replication Strategy

Replication to depression, anxiety, sleep, burnout, post-traumatic stress disorder, obsessive-compulsive disorder, bipolar disorder, autism, and peak performance proceeds by populating data rather than writing code. For each new symptom domain, the team curates a training library seed (a set of evidence-based interventions with full quick_input_schema, fischer_start_level, and contraindications fields), assembles a contraindication rule set, and validates the recommendation engine's outputs against expert clinical judgment on a representative case sample. The same six algorithms and the same database schema serve all domains. Domain-specific evidence sources include Cuijpers et al. (2020) for depression, Carpenter et al. (2018) for anxiety, Riemann et al. (2023) for sleep, Maslach and Leiter (2016) for burnout, Bisson et al. (2020) for post-traumatic stress disorder, Skapinakis et al. (2016) for obsessive-compulsive disorder, Goodwin et al. (2016) for bipolar disorder, and Lord et al. (2022) for autism.

13.3 Training Library Seed Data Curation

Per-symptom training library seeds are curated by a clinician with combined doctoral-level training, Board Certification in Neurofeedback, and Harvard Graduate School of Education affiliation (former Visiting Scholar). Seeds prioritize interventions with rct or meta_analysis evidence_level where available, with consensus_guideline and expert_opinion entries filling gaps. Each seed is reviewed for cultural appropriateness across Korean and English clinical practice (Hwang et al., 2008). The curated library is itself a form of intellectual property, distinct from but complementary to the algorithmic engine.

13.4 Cross-Symptom Comorbidity Handling

Comorbidity is the rule rather than the exception in real-world mental health populations. The unified profile schema accommodates comorbidity through the symptom_priorities vector, which encodes relative weighting across domains rather than a categorical single-symptom assignment. ALG-1 selects the primary symptom domain by argmax of symptom_priorities but uses the full vector to flag relevant secondary considerations. For example, a patient with primary ADHD and secondary depression may receive an ADHD training recommendation with explicit notation that depression-friendly modalities (behavioral activation overlap) are preferred. Full multi-symptom planning across simultaneous concurrent treatments is reserved for future work.

14. Discussion

14.1 Strengths

The architecture's principal strengths are integration, privacy, and traceability. By unifying assessment, recommendation, session planning, and adaptive replanning under one schema, the design removes the structural gap between diagnosis and treatment that characterizes most contemporary digital mental health interventions. By treating the bot-clinical bridge as an opt-in default-off explicit decision, the design respects patient autonomy while preserving the option to integrate when appropriate. By persisting full score decomposition, evidence citations, bilingual reasoning, and model version with every recommendation, the design provides regulators, clinicians, and patients with the substrate for meaningful audit and accountability.

A further strength is the universal symptom-domain structure. The same six algorithms and the same database schema serve all ten Catchers; new symptom domains are added by populating data rather than writing code. This minimizes implementation cost per domain and ensures consistency in audit, privacy handling, and bilingual rendering across the full platform.

14.2 Limitations

Several limitations apply. First, the engine has not yet been clinically validated; planned validation includes recommendation-alignment studies against expert clinician choice, reasoning-quality studies with external reviewers, and bilingual comprehensibility studies. Second, the cold-start problem persists: in the absence of prior outcome data, the personal_history component contributes a neutral prior, which may lead to recommendations that revert to evidence-strength dominance. Third, the engine assumes that clinician-curated content is correct; bias in the training_library seed propagates to recommendations. Fourth, Fischer skill estimation in real time is non-trivial; current implementation uses self-report and clinician-administered shorthand assessment, which may be insufficient for fine-grained level discrimination.

14.3 Clinical Implications

For practicing clinicians, the architecture offers a clinician-AI collaboration model rather than clinician replacement. ALG-1 produces ranked recommendations with reasoning; the clinician selects, modifies, or overrides. ALG-5 allows clinicians to author bespoke trainings that conform to the recommendation engine's expectations. The bot-bridge architecture allows clinicians to access bot-derived clinical signal only with explicit patient consent, preserving the therapeutic alliance. We hypothesize, though do not here demonstrate, that this collaboration model increases clinician acceptance over autonomous algorithmic recommendation systems.

14.4 Regulatory Pathway

The architecture is designed to support a U.S. Food and Drug Administration Software-as-a-Medical-Device Class II clearance pursuit. The audit traceability layer addresses Good Machine Learning Practice principles regarding transparency, post-market monitoring, and lifecycle audit. The opt-in default-off bot bridge addresses General Data Protection Regulation Article 9 and the European Union AI Act's high-risk system transparency requirements. The bilingual rendering pipeline supports international deployment. Conformity to either framework requires outcome validation across diverse populations, which is the primary aim of the ADHD Gold Template validation phase.

14.5 Future Work

Three lines of future work are prioritized. First, outcome data integration: as patients complete sessions and report outcomes, the engine should update its scoring through supervised learning while preserving the audit trail. Second, multi-modal biomarker expansion: integration of wearable physiological data (heart rate variability, sleep architecture from consumer devices) into biomarker_summary would strengthen the biomarker_match component. Third, multi-symptom concurrent treatment planning: real-world patients often warrant simultaneous attention to multiple symptom domains; extending ALG-2 to plan concurrent rather than sequential trainings is a substantial design problem.

15. Conclusion

We have presented an AI-driven personalized treatment pathway engine for multi-symptom digital mental health platforms. The architecture comprises six algorithms operating against a unified relational database schema: a multi-component recommendation engine, a Fischer-aware session plan generator, an adaptive replanner, a privacy-preserving bot-clinical bridge, a custom training builder, and a bilingual result renderer. The design addresses three persistent gaps in contemporary digital mental health: the disconnect between assessment and personalized treatment planning, the privacy paradox of bot-clinical integration, and the absence of audit-traceable reasoning for algorithmic recommendations.

Three architectural commitments distinguish this work. First, treatment pathway logic is explicit, modular, and auditable rather than implicit in unstructured chatbot conversation or static psychoeducational content. Second, the bridge between conversational and clinical data streams is governed by explicit opt-in default-off consent semantics rather than blanket integration. Third, the entire architecture is universal across ten symptom domains, with ADHD serving as a Gold Template for clinical validation before replication to depression, anxiety, sleep, burnout, post-

traumatic stress disorder, obsessive-compulsive disorder, bipolar disorder, autism, and peak performance.

The system is designed to be regulator-aligned with United States Food and Drug Administration Software-as-a-Medical-Device guidance, the European Union AI Act, the General Data Protection Regulation, and the Health Insurance Portability and Accountability Act. The architecture itself is necessary but not sufficient for regulatory clearance; outcome validation across diverse populations and post-market monitoring remain required prior to any clearance pursuit. The next phase of work is clinical validation of the ADHD Gold Template, after which replication to the remaining nine symptom domains can proceed as a primarily data-curation rather than software-engineering exercise.

Beyond its immediate platform context, the architecture offers a generalizable template for the construction of multi-symptom digital mental health platforms that integrate conversational and clinical components without sacrificing privacy or transparency. The Fischer dynamic skill compatibility component, in particular, may have applications beyond the present platform wherever developmentally calibrated treatment matching is required. We hope this work contributes to a digital mental health ecosystem in which personalization, privacy, and accountability are not in tension but are jointly designed from the beginning.

References

Almirall, D., Nahum-Shani, I., Sherwood, N. E., & Murphy, S. A. (2014). Introduction to SMART designs for the development of adaptive interventions: With application to weight loss research. Translational Behavioral Medicine, 4(3), 260–274. https://doi.org/10.1007/s13142-014-0265-0

Arns, M., de Ridder, S., Strehl, U., Breteler, M., & Coenen, A. (2009). Efficacy of neurofeedback treatment in ADHD: The effects on inattention, impulsivity and hyperactivity: A meta-analysis. Clinical EEG and Neuroscience, 40(3), 180–189. https://doi.org/10.1177/155005940904000311

Baumel, A., Muench, F., Edan, S., & Kane, J. M. (2019). Objective user engagement with mental health apps: Systematic search and panel-based usage analysis. Journal of Medical Internet Research, 21(9), e14567. https://doi.org/10.2196/14567

Bisson, J. I., van Gelderen, M., Roberts, N. P., & Lewis, C. (2020). Non-pharmacological and non-psychological approaches to the treatment of PTSD: Results of a systematic review and meta-analyses. European Journal of Psychotraumatology, 11(1), 1795361. https://doi.org/10.1080/20008198.2020.1795361

Carpenter, J. K., Andrews, L. A., Witcraft, S. M., Powers, M. B., Smits, J. A. J., & Hofmann, S. G. (2018). Cognitive behavioral therapy for anxiety and related disorders: A meta-analysis of randomized placebo-controlled trials. Depression and Anxiety, 35(6), 502–514. https://doi.org/10.1002/da.22728

Cohen, I. G., & Mello, M. M. (2018). HIPAA and protecting health information in the 21st century. JAMA, 320(3), 231–232. https://doi.org/10.1001/jama.2018.5630

Cortese, S., Adamo, N., Del Giovane, C., Mohr-Jensen, C., Hayes, A. J., Carucci, S., Atkinson, L. Z., Tessari, L., Banaschewski, T., Coghill, D., Hollis, C., Simonoff, E., Zuddas, A., Barbui, C., Purgato, M., Steinhausen, H.-C., Shokraneh, F., Xia, J., & Cipriani, A. (2018). Comparative efficacy and tolerability of medications for attention-deficit hyperactivity disorder in children, adolescents, and adults: A systematic review and network meta-analysis. The Lancet Psychiatry, 5(9), 727–738. https://doi.org/10.1016/S2215-0366(18)30269-4

Cuijpers, P., Karyotaki, E., Eckshtain, D., Ng, M. Y., Corteselli, K. A., Noma, H., Quero, S., & Weisz, J. R. (2020). Psychotherapy for depression across different age groups: A systematic review and meta-analysis. JAMA Psychiatry, 77(7), 694–702. https://doi.org/10.1001/jamapsychiatry.2020.0164

Darcy, A., Daniels, J., Salinger, D., Wicks, P., & Robinson, A. (2021). Evidence of human-level bonds established with a digital conversational agent: Cross-sectional, retrospective observational study. JMIR Formative Research, 5(5), e27868. https://doi.org/10.2196/27868

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608. https://arxiv.org/abs/1702.08608

Fischer, K. W. (1980). A theory of cognitive development: The control and construction of hierarchies of skills. Psychological Review, 87(6), 477–531. https://doi.org/10.1037/0033-295X.87.6.477

Fischer, K. W., & Bidell, T. R. (2006). Dynamic development of action and thought. In W. Damon & R. M. Lerner (Eds.), Handbook of child psychology: Theoretical models of human development (6th ed., Vol. 1, pp. 313–399). Wiley.

Fleming, T., Bavin, L., Lucassen, M., Stasiak, K., Hopkins, S., & Merry, S. (2018). Beyond the trial: Systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. Journal of Medical Internet Research, 20(6), e199. https://doi.org/10.2196/jmir.9275

Ghassemi, M., Oakden-Rayner, L., & Beam, A. L. (2021). The false hope of current approaches to explainable artificial intelligence in health care. The Lancet Digital Health, 3(11), e745–e750. https://doi.org/10.1016/S2589-7500(21)00208-9

Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a 'right to explanation'. AI Magazine, 38(3), 50–57. https://doi.org/10.1609/aimag.v38i3.2741

Goodwin, G. M., Haddad, P. M., Ferrier, I. N., Aronson, J. K., Barnes, T. R. H., Cipriani, A., Coghill, D. R., Fazel, S., Geddes, J. R., Grunze, H., Holmes, E. A., Howes, O., Hudson, S., Hunt, N., Jones, I., Macmillan, I. C., McAllister-Williams, H., Miklowitz, D. R., Morriss, R., … Young, A. H. (2016). Evidence-based guidelines for treating bipolar disorder: Revised third edition recommendations from the British Association for Psychopharmacology. Journal of Psychopharmacology, 30(6), 495–553. https://doi.org/10.1177/0269881116636545

Heinz, M. V., Mackin, D. M., Trudeau, B. M., Bhattacharya, S., Wang, Y., Banta, H. A., Jewett, A. D., Salzhauer, A. J., Griffin, T. Z., & Jacobson, N. C. (2025). Randomized trial of a generative AI chatbot for mental health treatment. NEJM AI, 2(4). https://doi.org/10.1056/AIoa2400802

Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds.) (2024). Cochrane Handbook for Systematic Reviews of Interventions (Version 6.5). Cochrane. https://training.cochrane.org/handbook

Hornstein, S., Zantvoort, K., Lueken, U., Funk, B., & Hilbert, K. (2023). Personalisation strategies in digital mental health interventions: A systematic review and conceptual framework for depressive symptoms. Frontiers in Digital Health, 5, 1170002. https://doi.org/10.3389/fdgth.2023.1170002

Hwang, W.-C., Myers, H. F., Abe-Kim, J., & Ting, J. Y. (2008). A conceptual paradigm for understanding culture's impact on mental health: The cultural influences on mental health (CIMH) model. Clinical Psychology Review, 28(2), 211–227. https://doi.org/10.1016/j.cpr.2007.05.001

Lee, K. H., Choi, G. H., Yun, J., Choi, J., Goh, M. J., Sinn, D. H., Jin, Y. J., Kim, M. A., Yu, S. J., Jang, S., Lee, S. K., Jang, J. W., Lee, J. S., Kim, D. Y., Cho, Y. Y., Kim, H. J., Kim, S., Kim, J. H., Kim, N., & Kim, K. M. (2024). Machine learning-based clinical decision support system for treatment

recommendation and overall survival prediction of hepatocellular carcinoma: A multi-center study. npj Digital Medicine, 7, 2. https://doi.org/10.1038/s41746-023-00976-8

Linardon, J., Cuijpers, P., Carlbring, P., Messer, M., & Fuller-Tyszkiewicz, M. (2024). The efficacy of app-supported smartphone interventions for mental health problems: A meta-analysis of randomized controlled trials. World Psychiatry, 23(1), 139–149. https://doi.org/10.1002/wps.21183

Lord, C., Charman, T., Havdahl, A., Carbone, P., Anagnostou, E., Boyd, B., Carr, T., de Vries, P. J., Dissanayake, C., Divan, G., Freitag, C. M., Gotelli, M. M., Kasari, C., Knapp, M., Mundy, P., Plank, A., Scahill, L., Servili, C., Shattuck, P., … McCauley, J. B. (2022). The Lancet Commission on the future of care and clinical research in autism. The Lancet, 399(10321), 271–334. https://doi.org/10.1016/S0140-6736(21)01541-5

Mascolo, M. F., & Fischer, K. W. (2015). Dynamic development of thinking, feeling, and acting. In W. F. Overton & P. C. M. Molenaar (Eds.), Handbook of child psychology and developmental science: Theory and method (7th ed., Vol. 1, pp. 113–161). Wiley. https://doi.org/10.1002/9781118963418.childpsy104

Maslach, C., & Leiter, M. P. (2016). Understanding the burnout experience: Recent research and its implications for psychiatry. World Psychiatry, 15(2), 103–111. https://doi.org/10.1002/wps.20311

Mohr, D. C., Zhang, M., & Schueller, S. M. (2017). Personal sensing: Understanding mental health using ubiquitous sensors and machine learning. Annual Review of Clinical Psychology, 13, 23–47. https://doi.org/10.1146/annurev-clinpsy-032816-044949

Murphy, S. A. (2005). An experimental design for the development of adaptive treatment strategies. Statistics in Medicine, 24(10), 1455–1481. https://doi.org/10.1002/sim.2022

Nahum-Shani, I., Smith, S. N., Spring, B. J., Collins, L. M., Witkiewitz, K., Tewari, A., & Murphy, S. A. (2018). Just-in-time adaptive interventions (JITAIs) in mobile health: Key components and design principles for ongoing health behavior support. Annals of Behavioral Medicine, 52(6), 446–462. https://doi.org/10.1007/s12160-016-9830-8

National Board for Certified Counselors (2023). National Clinical Mental Health Counseling Examination (NCMHCE). NBCC. https://www.nbcc.org/exams/ncmhce

Reyes-Portillo, J. A., Mufson, L., Greenhill, L. L., Gould, M. S., Fisher, P. W., Tarlow, N., & Rynn, M. A. (2025). Generative AI mental health chatbots as therapeutic tools: Systematic review and meta-analysis. JMIR Mental Health, 12, e69220. https://doi.org/10.2196/69220

Riemann, D., Espie, C. A., Altena, E., Arnardottir, E. S., Baglioni, C., Bassetti, C. L. A., Bastien, C., Berzina, N., Bjorvatn, B., Dikeos, D., Dolenc Groselj, L., Ellis, J. G., Garcia-Borreguero, D., Geoffroy, P. A., Gjerstad, M., Gonçalves, M., Hertenstein, E., Hoedlmoser, K., Hion, T., … Spiegelhalder, K. (2023). The European Insomnia Guideline: An update on the diagnosis and treatment of insomnia 2023. Journal of Sleep Research, 32(6), e14035. https://doi.org/10.1111/jsr.14035

Skapinakis, P., Caldwell, D. M., Hollingworth, W., Bryden, P., Fineberg, N. A., Salkovskis, P., Welton, N. J., Baxter, H., Kessler, D., Churchill, R., & Lewis, G. (2016). Pharmacological and psychotherapeutic interventions for management of obsessive-compulsive disorder in adults: A systematic review and network meta-analysis. The Lancet Psychiatry, 3(8), 730–739. https://doi.org/10.1016/S2215-0366(16)30069-4

Sutton, R. T., Pincock, D., Baumgart, D. C., Sadowski, D. C., Fedorak, R. N., & Kroeker, K. I. (2020). An overview of clinical decision support systems: Benefits, risks, and strategies for success. npj Digital Medicine, 3, 17. https://doi.org/10.1038/s41746-020-0221-y

Tjoa, E., & Guan, C. (2021). A survey on explainable artificial intelligence (XAI): Toward medical XAI. IEEE Transactions on Neural Networks and Learning Systems, 32(11), 4793–4813. https://doi.org/10.1109/TNNLS.2020.3027314

Torous, J., Bucci, S., Bell, I. H., Kessing, L. V., Faurholt-Jepsen, M., Whelan, P., Carvalho, A. F., Keshavan, M., Linardon, J., & Firth, J. (2021). The growing field of digital psychiatry: Current evidence and the future of apps, social media, chatbots, and virtual reality. World Psychiatry, 20(3), 318–335. https://doi.org/10.1002/wps.20883

U.S. Food and Drug Administration (2021). Artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD) action plan. U.S. Department of Health and Human Services. https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device

Van Doren, J., Arns, M., Heinrich, H., Vollebregt, M. A., Strehl, U., & Loo, S. K. (2019). Sustained effects of neurofeedback in ADHD: A systematic review and meta-analysis. European Child & Adolescent Psychiatry, 28(3), 293–305. https://doi.org/10.1007/s00787-018-1121-4

Voigt, P., & von dem Bussche, A. (2017). The EU General Data Protection Regulation (GDPR): A practical guide. Springer International Publishing. https://doi.org/10.1007/978-3-319-57959-7

Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. International Data Privacy Law, 7(2), 76–99. https://doi.org/10.1093/idpl/ipx005

Wang, Y., Li, X., Zhang, Q., Yeung, D., & Wu, Y. (2025). Effect of a cognitive behavioral therapy-based AI chatbot on depression and loneliness in Chinese university students: Randomized controlled trial with financial stress moderation. JMIR mHealth and uHealth, 13, e72365. https://doi.org/10.2196/72365

Wanniarachchi, V., Greenhalgh, M., Choi, A., & Warren, A. (2025). Personalization variables in digital mental health interventions for depression and anxiety in adolescents and youth: A scoping review. Frontiers in Digital Health, 7, 1500220. https://doi.org/10.3389/fdgth.2025.1500220

Warraich, H. J., Tazbaz, T., & Califf, R. M. (2024). FDA perspective on the regulation of artificial intelligence in health care and biomedicine. JAMA, 333(3), 241–247. https://doi.org/10.1001/jama.2024.21451

Wiens, J., Saria, S., Sendak, M., Ghassemi, M., Liu, V. X., Doshi-Velez, F., Jung, K., Heller, K., Kale, D., Saeed, M., Ossorio, P. N., Thadaney-Israni, S., & Goldenberg, A. (2019). Do no harm: A roadmap for responsible machine learning for health care. Nature Medicine, 25(9), 1337–1340. https://doi.org/10.1038/s41591-019-0548-6