

Co-design and development of an occupational therapy conversational AI system for neurodevelopmental care in Greece: A mixed-methods protocol

Pantelis Pergantis^{1,2}, Nikolaos Bardis³, Charalabos Skianis² & Athanasios Drigas¹

¹ Net Media Lab & Mind & Brain R&D, Institute of Informatics & Telecommunications, National Centre of Scientific Research 'Demokritos', 153 41 Agia Paraskevi, Greece

² Department of Information & Communication Systems Engineering, University of the Aegean, 832 00 Karlovasi, Greece

³ Mathematics and Engineering Science Sector, Hellenic Army Academy, 166 73 Vari, Greece

Correspondence: Pantelis Pergantis, Net Media Lab & Mind & Brain R&D, Institute of Informatics & Telecommunications, National Centre of Scientific Research 'Demokritos', 153 41 Agia Paraskevi, Greece; Department of Information & Communication Systems Engineering, University of the Aegean, 832 00 Karlovasi, Greece. E-mail: p.pergantis@iit.demokritos.gr

Submitted: June 15, 2026

DOI: 10.14295/bjs.v6i2.943

Accepted: July 27, 2026

URL: <https://doi.org/10.14295/bjs.v6i2.943>

Published: July 28, 2026

Abstract

Conversational artificial intelligence (AI) has shown potential to support knowledge translation, personalized education, and access to health-related information. However, applications in neurodevelopmental care remain largely focused on screening and assessment, while conversational systems tailored to occupational therapy are scarce. Moreover, general-purpose generative AI may produce inaccurate, insufficiently contextualized, or clinically inappropriate responses, highlighting the need for evidence-based, domain-specific systems with robust safety mechanisms. This study describes the protocol for the co-design, development, and preliminary evaluation of a conversational AI system designed to support occupational therapists and parents of children aged 5–12 years with neurodevelopmental disorders in Greece. The system is intended as an educational and decision-support resource and will not replace diagnosis, clinical judgment, or individualized intervention. A co-designed, multi-phase, mixed-methods proof-of-concept design will be adopted, informed by the Medical Research Council framework, Design Science Research, user-centered design principles, and the CeHRes Roadmap 2.0. Development will include evidence synthesis, stakeholder needs assessment, knowledge base construction, iterative prototype development, and expert, technical, safety, and user evaluation. The system will integrate a curated occupational therapy knowledge base, retrieval-augmented generation, role-specific prompting, source verification, and layered safety guardrails. Expected outputs include a stakeholder-informed Greek-language minimum viable product and a transparent framework linking evidence, user requirements, technical design, and evaluation criteria. Preliminary evaluation will assess factual accuracy, evidence concordance, occupational therapy relevance, clinical appropriateness, safety, usability, acceptability, and perceived usefulness. This protocol provides a reproducible foundation for developing clinically relevant conversational AI in occupational therapy and for future feasibility and effectiveness studies.

Keywords: artificial intelligence, conversational AI, chatbot, occupational therapy, neurodevelopmental disorders.

Cocriação e desenvolvimento de um sistema de IA conversacional em terapia ocupacional para cuidados em neurodesenvolvimento na Grécia: Protocolo de métodos mistos

Resumo

A inteligência artificial (IA) conversacional tem demonstrado potencial para apoiar a tradução do conhecimento, a educação personalizada e o acesso a informações relacionadas à saúde. No entanto, suas aplicações no cuidado aos transtornos do neurodesenvolvimento permanecem amplamente concentradas na triagem e avaliação, enquanto sistemas conversacionais voltados especificamente para a terapia ocupacional ainda são escassos. Além disso, sistemas generativos de uso geral podem produzir respostas imprecisas, insuficientemente contextualizadas ou clinicamente inadequadas, destacando a necessidade de soluções baseadas em evidências e dotadas de mecanismos robustos de segurança. Este estudo descreve o protocolo para o codesenvolvimento, a construção e a avaliação preliminar de um sistema de IA conversacional destinado a apoiar terapeutas ocupacionais e pais de crianças de 5 a 12 anos com transtornos do neurodesenvolvimento na Grécia. O sistema foi concebido como um recurso educacional e de apoio à tomada de decisão, não substituindo o diagnóstico, o julgamento clínico ou as intervenções individualizadas. Será adotado um delineamento de prova de conceito, em múltiplas fases, com abordagem de métodos mistos e desenvolvimento colaborativo, fundamentado no framework do Medical Research Council, na Design Science Research, nos princípios do design centrado no usuário e no CeHRes Roadmap 2.0. O desenvolvimento incluirá síntese de evidências, levantamento das necessidades dos stakeholders, construção da base de conhecimento, desenvolvimento iterativo do protótipo e avaliações por especialistas, além de avaliações técnicas, de segurança e com usuários. O sistema integrará uma base de conhecimento curada em terapia ocupacional, geração aumentada por recuperação (Retrieval-Augmented Generation – RAG), arquitetura de prompts específica para cada perfil de usuário, verificação de fontes e múltiplas camadas de proteção. Espera-se produzir um produto mínimo viável em língua grega, desenvolvido com participação dos stakeholders, além de um framework transparente que conecte evidências, requisitos dos usuários, decisões técnicas e critérios de avaliação. A avaliação preliminar contemplará precisão factual, concordância com as evidências, relevância para a terapia ocupacional, adequação clínica, segurança, usabilidade, aceitabilidade e utilidade percebida. Este protocolo estabelece uma base metodológica e técnica reproduzível para o desenvolvimento de IA conversacional clinicamente relevante em terapia ocupacional e para futuros estudos de viabilidade e efetividade.

Palavras-chave: inteligência artificial, IA conversacional, chatbot, terapia ocupacional, transtornos do neurodesenvolvimento.

1. Introduction

Artificial Intelligence (AI) is influencing more and more health care, rehabilitation, education, and the way professionals and the public access and interpret health-related information. Conversational agents, including rule-based chatbots and newer systems that are powered by Large Language models (LLMs), give information via an interactive natural-language interface. Many systematic reviews have found applications in patient education, health-behaviour support, clinical communication, and self-management, while giving emphasis to heterogeneous evidence, limited reproducibility, and many concerns regarding the accuracy, safety, transparency, and bias (Busch et al., 2025; Milne-Ives et al., 2020).

The potential value of conversational AI may be particularly relevant within neurodevelopmental care. Neurodevelopmental disorders comprise a heterogeneous group of conditions associated with differences or difficulties in cognitive, motor, communicative, behavioural, sensory, emotional, and adaptive functioning. Families may need to coordinate health, education, and social-care services, interpret complex professional recommendations, and adapt strategies to everyday routines. Research with parents of children with neurodevelopmental disorders has highlighted the need for digital resources that are understandable, trustworthy, relevant, and responsive to the priorities of intended users (Bui et al., 2022).

Many studies have started to investigate conversational technologies for neurodevelopmental populations and their families. Singla et al. (2024) described the development of the Coaching Assistant for Medical/Health Information, a domain-specific chatbot designed to connect families with trusted information and services. Its architecture incorporated a knowledge graph, professional input, standard and lay terminology, and consultation with people with lived experience. This study showed that specialized information can be organized within a conversational interface and also by identifying practical challenges that involve the coverage of terminology, entity recognition, and interpretation of users' natural language.

A related user-centred study showed that chatbot features cannot be selected solely based on developer assumptions. Parents expressed differing views related to customisation, rewards, social interaction, and restricted content, and they raised concerns about information accuracy (Bui et al., 2022). All these findings support continuous stakeholder involvement during system design rather than exclusively technology-driven

development.

Beyond information retrieval, conversational AI may have relevance for cognitive, metacognitive, and self-regulatory support. Pergantis et al. (2025a) found new data on working memory, attention, cognitive flexibility, decision-making, problem-solving, self-control, and metacognitive processes. The review offered a theoretical framework for examining structured conversational feedback, reflection, planning, and adaptive guidance in future systems, but it did not prove therapeutic efficacy.

A wider use of assistive and emerging technologies in neurodevelopmental care has been shown by related research. Pergantis et al. (2025b) examined methods for identifying or lowering neurophysiological stress and anxiety in children and adolescents with autism and sensory processing issues. Wearable technology, biofeedback, virtual and augmented environments, robotics, mobile systems, AI-supported detection, and serious games were among the technologies. These findings are relevant to sensory and emotional regulation, environmental adaptation, routines, and caregiver education, but they do not establish that a conversational chatbot can independently assess or treat stress.

AI research in Developmental Coordination Disorder shifted toward screening, classification, early identification, movement analysis, and assessment. Pergantis et al. (2026) identified comparatively limited evidence concerning AI-assisted intervention and rehabilitation and recommended future investigation of clinician-centred, theory-informed systems, including LLMs and conversational AI for clinical education, reasoning, and personalised support. Although Developmental Coordination Disorder represents only one diagnostic group, these findings illustrate a broader translational gap between AI-enabled identification and participation-oriented support.

A wider review of pediatric rehabilitation showed that AI-supported interventions were frequently aimed at robotics and human-machine interaction, whereas remote delivery, individual goal setting and adaptation to family-reported participation needs were not common (Kaelin et al., 2021). This provides a stronger case for the need for AI systems that are deliberately geared towards occupational participation, family needs and contextualised support, rather than just impairment classification.

Early work with parents suggests that this is possible, but also that its scope is limited. Zhang et al. (2026) assessed a GPT-powered positive-psychology chatbot with parents of autistic children and reported favourable feasibility and acceptability during a brief intervention period. However, the system mainly provided structured positive-psychology exercises, instead of occupational therapy-focused guidance or differentiated support for both parents and professionals.

Occupational therapy is particularly relevant in that it studies the interaction between people, their occupations, environmental conditions and contextual demands. Pediatric occupational therapists support self-care, play, education, social participation, family routines, motor performance, executive functioning, sensory processing, emotional regulation and family-centred goals. However, the use of AI in occupational therapy is still limited to areas such as motor rehabilitation, functional assessment, assistive technology, robotics and telerehabilitation (Bulan et al., 2025).

It cannot be assumed that general-purpose conversational systems replicate the context-sensitive clinical reasoning of occupational therapists. Değerli et al. (2026) compared the recommendations generated by ChatGPT with the decisions of 111 occupational therapists on uniform cases. Agreement was found in some areas, but significant differences emerged in others, highlighting the importance of person-specific, contextual, occupational, and environmental reasoning. These findings support the development of an occupational therapy-focused system rather than unrestricted use of a general-purpose chatbot.

The proposed study also grows out of a developing national research context. The research on occupational therapy in connection with Greek higher education institutions was limited for several decades but increased substantially after 2021 (Onoufriou et al. 2027). The analysis revealed fragmented collaboration and limited visibility, and also developing thematic areas such as neurodevelopment, cognitive functioning, rehabilitation and emerging technologies. These national findings do not by themselves support an international research gap, but they do provide context for the need for organized AI-related occupational therapy research in Greece.

Because LLM output is authoritative and fluid, users might overestimate reliability. Biased content, poor transparency, privacy concerns, ambiguous accountability, and reliable but risky health advice have all been found in reviews (Busch et al., 2025; Haltaufderheide; Ranisch, 2024). Consequently, a health-related conversational system should be built to identify situations that require professional referral or refusal, be based on carefully chosen evidence, have a clearly defined scope, be transparently sourced, and be continuously

monitored by humans.

Retrieval-augmented generation (RAG) provides one technical approach for connecting LLM output with a controlled external knowledge source. RAG combines the language-generation capacity of a pretrained model with passages retrieved from an indexed corpus, potentially improving evidence grounding and facilitating knowledge updates without retraining the underlying model (Lewis et al., 2020). However, retrieval errors, unsupported synthesis, citation mismatches, and hallucinations remain possible, requiring systematic technical and expert evaluation.

The international literature increasingly supports co-design and iterative development. Livermore et al. (2024) provided a directly relevant protocol for a co-designed proof-of-concept chatbot for parents and caregivers of children and young people with rheumatological conditions. Their staged approach included needs exploration, stakeholder workshops, minimum viable product development, and early user testing before progression to effectiveness evaluation. The CeHRes Roadmap 2.0 similarly emphasises contextual inquiry, value specification, iterative design, operationalisation, and continuous formative evaluation (Kip et al., 2025).

The literature identified to date does not appear to describe an evidence-based conversational AI system that systematically combines an occupational therapy framework with separate, role-specific support for occupational therapists and parents of children with neurodevelopmental disorders. The systems that already exist primarily focus on information retrieval, the wellbeing of the parents as well as general health education or condition-specific assistance.

Therefore, the current study aims to describe the co-design, technical development, and planned preliminary evaluation of an evidence-based, occupational therapy-focused conversational AI system for occupational therapists and parents of children aged 5–12 years with neurodevelopmental disorders. The system will be developed as an educational and decision-support resource rather than an autonomous diagnostic or therapeutic tool. The protocol will document evidence acquisition, stakeholder value specification, knowledge architecture, model configuration, RAG, prompt engineering, safety guardrails, technical testing, and planned expert and user evaluation.

2. Materials and Methods

2.1 Study design and methodological frameworks

This study will adopt a co-designed, multi-phase, mixed-methods experimental proof-of-concept design to guide the systematic development and formative evaluation of an evidence-based conversational AI system grounded in occupational therapy. The system will be designed for occupational therapists and parents or primary caregivers of children aged 5–12 years with neurodevelopmental disorders.

The first minimum viable product will target children between the ages of five and twelve with autism spectrum disorder, attention-deficit/hyperactivity disorder, and developmental coordination disorder. The knowledge architecture will be limited to five main occupational and functional domains within these diagnostic groups: self-care and daily routines; play, education, and social participation; executive functioning and self-regulation; motor coordination and praxis; and sensory and emotional regulation. The prototype will not attempt to address every clinical presentation, intervention approach, comorbidity, or occupational need associated with these conditions. Additional neurodevelopmental conditions and knowledge domains will be considered only in later versions after the initial system has demonstrated acceptable technical performance, safety, occupational therapy relevance, usability, and feasibility.

The study will focus on the creation and formative evaluation of the system. It will not evaluate diagnosis, medication decisions, clinical effectiveness, or changes in child-level occupational outcomes. Children will not participate directly, and the system will not be tested through access to clinical records or autonomous clinical decision-making.

Five interconnected phases will be undertaken: evidence synthesis and contextual inquiry; stakeholder needs assessment and value specification; knowledge-base construction and system-requirement development; iterative co-design and prototype engineering; and expert, technical, safety, and preliminary user evaluation. Although reported sequentially, the process will remain iterative.



Figure 1. Planned multi-phase development and evaluation pathway. Source: Authors, 2026.

The CeHRes Roadmap 2.0 will serve as the principal framework guiding the development process. This framework conceptualises the development, implementation, and evaluation of digital-health technologies as interconnected and iterative processes that require sustained stakeholder participation, interdisciplinary collaboration, continuous formative evaluation, and careful consideration of the relationship between technology, users, and the context of implementation (Kip et al., 2025).

Within the present study, the principles of contextual inquiry, value specification, design, operationalisation, and evaluation will be adapted to the development of the proposed conversational AI system. In this study, we want to highlight the need to ensure contextual fit, identify values important to occupational therapists and parents, document the relationship between stakeholder needs and system features, and progressively refine the prototype through repeated evaluation.

Further methodological guidance will be provided by the revised Medical Research Council framework for creating and assessing complex interventions. The suggested conversational AI system will be viewed as a sophisticated digital intervention whose operation and results could be impacted by its intended users, implementation context, technical architecture, and interaction with professional practice.

The current protocol mostly aligns with the framework's development and feasibility phases. Until the system has shown an acceptable degree of factual accuracy, occupational therapy relevance, safety, usability, and acceptability, clinical efficacy and large-scale implementation will not be assessed (Skivington et al., 2021).

The principles of Design Science Research will serve as a guide for the development of the conversational system as a technological artifact. This method comprises identifying a relevant problem, describing the objectives of a potential solution, designing and producing the artifact, demonstrating its intended uses, evaluating its efficacy, and disseminating the knowledge generated during the development process (Peppers et al., 2007).

Additionally, by facilitating the documentation of technical choices, specifications, prototype iterations, and evaluation processes, Design Science Research will assist the health-specific frameworks.

The intended contribution will therefore include both the final digital artifact and the methodological knowledge produced by its design and enhancement.

The study will include co-design and user-centered development. Occupational therapists and parents will participate in the stages of prototype development, value specification, and contextual inquiry. Their involvement will guarantee that the system's features, terminology, response style, boundaries, and interaction pathways represent actual user needs rather than conjectures made solely by technical developers or researchers.

Stakeholder consultation, co-design workshops, minimum viable product development, and iterative user testing before more thorough effectiveness studies are all consistent with indexed chatbot-development protocols (Livermore et al., 2024).

2.2 Study setting, participants, eligibility, and recruitment

The study will be conducted in Greece between September 2026 and December 2027. Stakeholder consultations, co-design sessions, expert evaluations, and preliminary user testing will be conducted primarily online. This format will facilitate the participation of occupational therapists and parents from different geographical regions of the country and will reduce the practical demands associated with in-person attendance.

The initial prototype will operate in Greek. Its terminology, response style, degree of detail, and conversational structure will be adapted according to whether the user is an occupational therapist or a parent. Scientific sources included in the knowledge base may be published in English, but relevant information will be communicated to users in Greek.

The study will be coordinated by Net Media Lab & Mind & Brain R&D, Institute of Informatics and Telecommunications, National Centre for Scientific Research "Demokritos", Greece, in collaboration with the

Department of Information and Communication Systems Engineering, University of the Aegean. The interdisciplinary research team will include expertise in occupational therapy, neurodevelopmental disorders, executive functioning, digital health, artificial intelligence, and research methodology.

Three groups will contribute to different stages of the study. The stakeholder co-design group will include approximately 8–10 occupational therapists and 8–10 parents or primary caregivers. Their contribution will focus on identifying user needs, values, expectations, concerns, and preferred system features. This sample is intended to support formative development rather than statistical generalisation.

The system will be assessed by an interdisciplinary panel of five to seven experts. The experts will mostly be professionals with experience in neurodevelopmental disorders and pediatric occupational therapists. Whenever feasible, at least one member with experience in digital health or artificial intelligence will be included. If a suitable expert is available, knowledge of digital ethics, health-data governance, or safety will also be represented. Each panelist may offer knowledge in more than one field.

Ten to twelve occupational therapists and ten to twelve parents or primary caregivers will participate in preliminary user testing. Participants in user testing will, whenever possible, be different from those who took part in the earlier stages of co-design. If recruitment restrictions make overlap inevitable, this will be noted and taken into account when interpreting the usability and acceptability findings.

A formal statistical power calculation will not be conducted because the study is formative and is not designed to test hypotheses concerning clinical effectiveness. The limited scope of the prototype, the goal of identifying primary design and usability issues, and the pragmatic demands of an early-stage proof-of-concept study are all reflected in the suggested sample sizes.

Participants will be occupational therapists who have at least a year of experience working with children or adolescents, have a recognized occupational therapy qualification, are legally allowed to practice in Greece, and have relevant academic or research experience. Additionally, they must be able to communicate in Greek, have experience with neurodevelopmental disorders, and give informed consent. Because their assessment of usability may be influenced by their prior knowledge of the system, occupational therapists who are directly involved in programming or technical development of the system will not take part in the user evaluation.

Parents and primary caregivers who are 18 years of age or older and who are in charge of a child between the ages of 5 and 12 who has a neurodevelopmental disorder covered by the system's initial scope are eligible. They must be able to express themselves in Greek and give their informed consent. It will not be necessary for parents to submit clinical reports, diagnostic records, or personally identifiable information about their child. Self-report will be used to determine eligibility.

In addition to having at least three years of clinical, academic, research, or technological experience in their field of expertise, members of the expert panel must have a relevant professional or academic qualification. Their experience must be relevant to paediatric occupational therapy, neurodevelopment, artificial intelligence, digital health, health-data governance, or digital ethics. Experts will also be required to disclose any potential conflicts of interest related to the selected technological platform or its provider.

Participants will be recruited through purposive and convenience-sampling procedures. Recruitment channels may include professional occupational therapy networks, universities, research centres, paediatric clinical services, parent organisations, professional social-media groups, and communication channels associated with the participating research laboratory.

Recruitment materials will provide a concise description of the study, its eligibility requirements, the anticipated duration of participation, the voluntary nature of involvement, and the procedures used to protect personal data. Contact details for the research team will also be provided. Participants will not be asked to disclose identifiable health information concerning a child during recruitment or data collection (Table 1).

Table 1. Planned study phases, contributors, and outputs.

Phase	Principal activities	Contributors	Primary outputs
1	Evidence update, contextual inquiry, and mapping of clinical and occupational domains	Research team, occupational therapists, parents	Evidence map and contextual requirements
2	Value specification and	Occupational	Requirements-traceability matrix

	conversion of needs into clinical, ethical, technical, and safety requirements	therapists, parents, research and technical teams	
3	Knowledge-base construction, document processing, retrieval architecture, prompt development, and safety design	Research and technical teams with expert review	Functional minimum viable product
4	Iterative co-design and refinement of language, interface, response structure, source display, and boundaries	Occupational therapists and parents	Revised stakeholder-informed prototype
5	Expert, technical, safety, and preliminary user evaluation	Experts, occupational therapists, parents	Formative evidence concerning technical performance, safety, occupational therapy relevance, usability, and acceptability

Source: Authors, 2026.

2.3 Phase 1: Evidence curation and contextual inquiry

A systematic process for identifying, evaluating, organizing, and maintaining evidence pertinent to the conversational AI system's intended scope will be used to build the scientific knowledge base. Since the research team has already looked at several pertinent domains in earlier systematic and scoping reviews, the procedure won't be a new stand-alone systematic review. To minimize selective source inclusion and guarantee that the data integrated into the system is up-to-date, pertinent, and methodologically sound, however, transparent and repeatable procedures will be used.

The initial evidence base will be informed by international evidence-based clinical guidelines, professional occupational therapy frameworks, systematic reviews, meta-analyses, scoping reviews addressing emerging areas, and high-quality primary studies where review-level evidence is unavailable. Ethical, technical, digital-health, data-governance, and AI-reporting guidance will also be included where directly relevant to the development or safe operation of the system.

Previous research examining AI chatbots and executive functioning will inform content related to cognitive control, planning, self-regulation, problem-solving, and metacognitive support (Pergantis et al., 2025a). Research concerning assistive and emerging technologies, autism, sensory-processing difficulties, stress, and anxiety will inform content related to sensory and emotional regulation and will contribute to the specification of the system's clinical boundaries (Pergantis et al., 2025b). The scoping review of artificial intelligence in Developmental Coordination Disorder will inform the development of participation-focused and intervention-related content (Pergantis et al., 2026). The bibliometric analysis of occupational therapy research in Greece will provide contextual information regarding the development of occupational therapy and digital-health research within the national setting (Onoufriou et al., 2027).

Before the final knowledge base is constructed, a targeted evidence update will be conducted in PubMed/MEDLINE, Scopus, Web of Science, CINAHL, PsycINFO, and IEEE Xplore. Searches will cover evidence published from the final search date of the preceding reviews until the date of the evidence update. Search strategies will combine terms related to occupational therapy, neurodevelopmental disorders, autism spectrum disorder, attention-deficit/hyperactivity disorder, Developmental Coordination Disorder, executive functioning, sensory and emotional regulation, occupational participation, conversational AI, chatbots, large language models, digital health, and caregiver or professional support.

The complete search strategies, dates of searching, databases, limits, and number of records identified will be provided in supplementary material. Searches will primarily include publications in English. Greek-language professional guidance may also be included where it originates from an identifiable professional, governmental, academic, or regulatory authority. Sources in other languages will be considered only when an accurate translation can be obtained and independently checked.

Sources will be accepted if they directly contribute to the ethical, technical, or safety architecture of the system, address at least one predetermined knowledge domain, involve a population or context pertinent to children with neurodevelopmental disorders, or offer information relevant to occupational participation or family and

professional support. Publications that offer commercial or promotional health advice, contain unverifiable information, are outside the prototype's defined population, professional, technical, or clinical scope, are purely opinion-based without supporting data, or lack an identifiable author or organization will be excluded.

Titles and abstracts identified through the evidence update will be screened by one member of the research team and independently checked by a second researcher. Potentially eligible full texts will be reviewed against the predefined eligibility criteria. Disagreements concerning inclusion will be resolved through discussion and, when necessary, consultation with a third researcher or an appropriate clinical or technical expert. A record will be maintained of the reasons for excluding full-text sources considered particularly relevant to the knowledge architecture.

A predetermined evidence hierarchy will be used to prioritize the included sources. Systematic reviews and meta-analyses will be given less weight than current evidence-based clinical guidelines and professional standards. Scoping reviews will be used to describe new fields or areas with inadequate evidence of efficacy. When review-level evidence is not available, is out of date, or is not sufficiently specific to the intended knowledge module, high-quality primary studies will be included.

The methodological quality and applicability of included sources will be considered using appraisal criteria appropriate to the study design. Clinical guidelines will be examined in relation to methodological transparency, recency, professional authority, and relevance to the intended population. Systematic and scoping reviews will be examined in relation to search comprehensiveness, eligibility criteria, appraisal procedures, synthesis methods, and limitations. Primary studies will be considered in relation to design, sample characteristics, risk of bias, outcome relevance, and consistency with the wider evidence base.

The purpose of quality appraisal will not be to generate a single pooled quality score. Instead, appraisal findings will be used to determine the confidence, emphasis, and language with which evidence is represented within the chatbot. Evidence derived from preliminary, indirect, heterogeneous, or low-certainty studies will not be presented with the same level of certainty as information supported by current guidelines or consistent review-level evidence.

When sources provide conflicting findings or recommendations, the research team will not select one position without documentation. The nature of disagreement, differences in population, intervention, context, methodology, and publication date will be examined. Where a defensible conclusion cannot be reached, the knowledge base will explicitly preserve the uncertainty and the chatbot will be instructed to communicate that the evidence is mixed, incomplete, or context-dependent.

The research team will document the full bibliographic citation, publication year, source type, population, neurodevelopmental condition, occupational domain, principal findings, methodological limitations, level of evidence, intended system module, date of inclusion, and review date for each included source. Every source will be assigned a unique identifier that will be associated with every text segment that comes from that particular document.

Evidence in English that is integrated into the Greek language system will go through a methodical translation and adaptation process. A bilingual member of the research team or a controlled translation tool may be used to prepare preliminary translations or Greek-language summaries. A Greek-speaking occupational therapist with pertinent professional experience will then review all translated clinical and occupational therapy content. Additionally, terminology meant for parents will be reviewed for clarity, accessibility, and avoidance of inappropriate and unprofessional language.

The evidence base will be treated as a version-controlled research resource. A stable version will be frozen before formal technical, expert, and user evaluation. During the evaluation period, no source will be added, deleted, or significantly altered without supporting documentation. Any necessary changes will be documented in the development log and result in a new version of the knowledge base.

Each source will be given a review date. When a relevant guideline is updated, withdrawn, or replaced; when significant new evidence becomes available; when expert evaluation reveals content that is out-of-date or inaccurate; or when a source is discovered to be at odds with current professional or regulatory guidance, it will be reevaluated. Material that has been withdrawn, superseded, or found to be methodologically flawed will be kept in the version history for auditing purposes but removed from active retrieval.

Separate online focus groups will be conducted with occupational therapists and parents. Each focus group will include approximately 5–8 participants and will last approximately 60–75 min. Conducting separate groups will allow participants to discuss their needs and concerns within a context relevant to their role.

A semi-structured discussion guide will examine frequently encountered information needs, common questions relating to executive functioning, sensory and emotional regulation, routines, participation, and intervention, and difficulties in locating or interpreting reliable information. Participants will also be invited to discuss their expectations of a conversational AI system, preferred terminology and response formats, concerns regarding accuracy and privacy, unacceptable system functions, circumstances in which professional referral should be recommended, and differences between the informational needs of occupational therapists and parents.

Focus groups will be audio-recorded following informed consent and transcribed verbatim. Any personally identifiable information or identifiable descriptions of children will be removed during transcription.

2.4 Phase 2: Value specification and system requirements

Findings from the evidence synthesis and contextual inquiry will be translated into explicit user, clinical, ethical, technical, accessibility, and safety requirements. A requirements-traceability matrix will be developed to document the relationship between each identified need and the corresponding system feature.

Each requirement's originating evidence or stakeholder concern, related user value, suggested technical or conversational solution, potential safety implications, method of verification, and final inclusion decision will all be documented in the matrix.

For instance, a requirement that parent-facing responses refrain from using superfluous clinical terminology may be derived from a parental need for clear information. Source-based answers and evidence summaries may be included due to a professional need for transparency. Diagnostic requests will be connected to professional referral pathways and predetermined boundary statements. Carefully crafted clarifying questions or clear statements about the limitations of the available information will be used to address situations involving inadequate contextual information.

Stakeholders will rate proposed requirements using a 5-point importance scale. A non-safety-related requirement will normally be retained when it receives a median score of at least 4 and when at least 70% of respondents rate it as important or very important. Safety-critical requirements may be retained regardless of stakeholder ratings when ethical, legal, or clinical considerations justify their inclusion.

2.5 Phase 3: Technical architecture and system development

A user interface layer, an application and orchestration layer, a knowledge and retrieval layer, a generative model layer, and a safety and verification layer will make up the secure web-based conversational application that will be created as the chatbot. Because of its modular design, it will be possible to test or replace individual components without having to rebuild the entire application.

Greek-language parent and occupational therapist pathways will be supported by the user interface. Role selection, query preprocessing, risk classification, retrieval, prompt assembly, model communication, post-generation checks, logging, and source presentation will all be managed by the orchestration layer. The knowledge layer will contain the curated scientific corpus and its semantic index. The generative layer will use a pre-existing LLM. The safety layer will operate before, during, and after response generation.

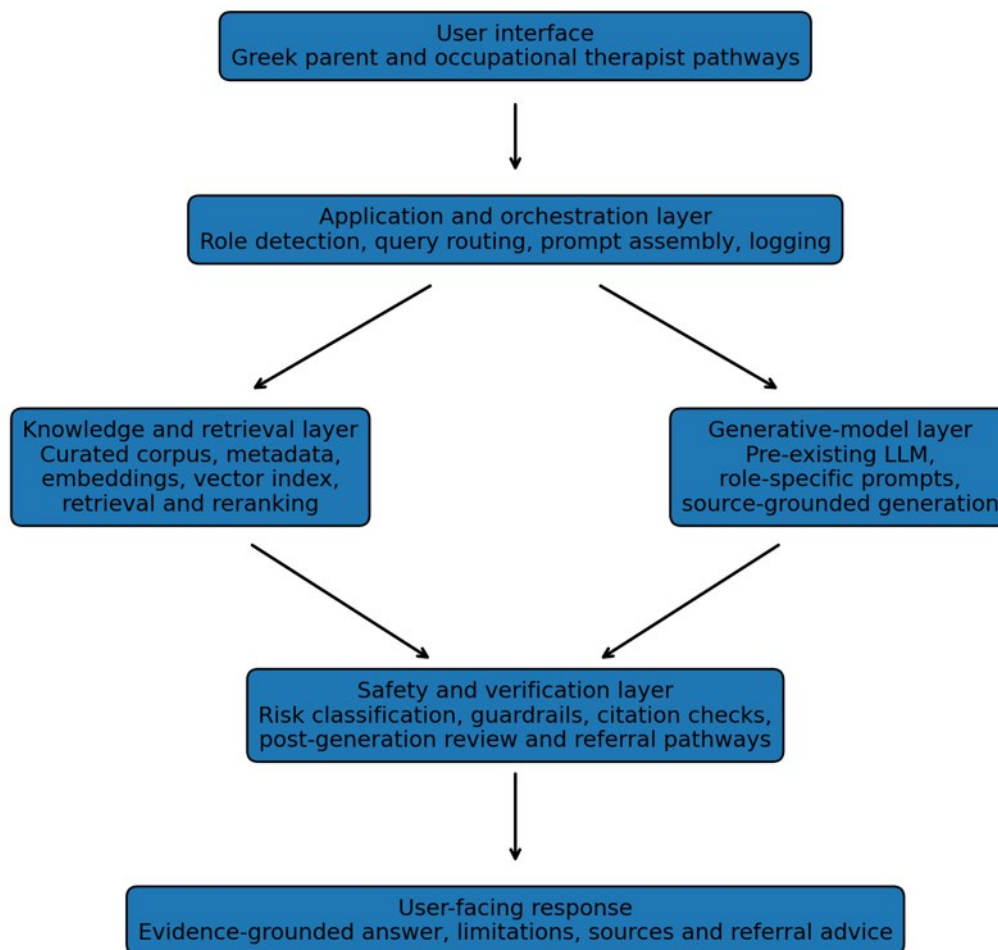


Figure 2. Planned modular technical architecture of the conversational AI system. Source: Authors, 2026.

The study will not train a foundational model from the beginning. Candidate pre-existing LLMs will be compared in relation to Greek-language performance, context-window capacity, response consistency, instruction following, technical accessibility, privacy and data-retention arrangements, cost, sustainability, and RAG compatibility. A small internal benchmark of representative Greek parent and professional questions will be used before final model selection.

The final report will identify the provider, exact model name and version, access route, dates of evaluation, context-window use, temperature, maximum output length, and other relevant generation settings. A stable model version and application release will be used throughout formal evaluation wherever technically possible.

The exact LLM, provider, model version, access route, context-window configuration, temperature, maximum output length, and other generation parameters will be selected through the predefined comparative procedure. Before formal evaluation, all chosen parameters and components will be recorded and frozen. Any further modifications will create a new version of the system and be noted in the development log.

A standardized ingestion process will be applied to eligible documents. Relevant navigation text, duplicate content, publisher material, and unrelated references will be eliminated, source files and metadata will be noted, and text will be extracted and examined for errors. Semantically coherent segments will be created from the cleaned evidence in order to preserve all recommendations, definitions, or conclusions. Only documents that have completed the predefined evidence-selection, appraisal, and approval procedure will be eligible for ingestion into the active knowledge base.

Metadata describing the source, evidence type, population, condition, occupational domain, intended user role, evidence priority, and review date will all be retained for each segment. All of this metadata will support the updating or removal of evidence in the future and make role-specific and topic-specific retrieval easier.

A multilingual sentence-embedding model that can process both Greek and English will be used to convert text segments into numerical semantic representations. Even when the user's wording is different from the source's, sentence embeddings enable the retrieval system to find conceptually related passages (Reimers; Gurevych, 2019). Cross-lingual performance will be crucial since many sources will be in English, but user queries will be written in Greek (Feng et al., 2022).

Since many sources will be in English but user queries will be written in Greek, cross-lingual performance will be essential (Feng et al., 2022). Greek queries connected to manually selected pertinent passages will be used to test candidate embedding models. A vector database or similarity-search index will be used to store vectors and metadata. The final report will include documentation of the precise embedding model, chunk size, overlap, indexing configuration, similarity metric, and vector-storage technology.

Greek-language queries connected to manually identified reference passages will be used for comparative testing to determine the final embedding model, document-chunking strategy, vector-storage technology, similarity metric, retrieval threshold, top-k value, and reranking configuration. Before formal evaluation, the entire configuration will be frozen and reported in the results paper and technical documentation that follow.

Role selection will be the first step in every interaction. After that, the input will be checked for high-risk subjects, out-of-scope requests, directly identifiable information, empty submissions, and inadequate context. The original meaning of eligible questions can be preserved by internally reformulating them into a succinct retrieval query.

Indexed evidence will be compared with the embedded query. Semantic similarity, keyword searching, metadata filtering, and reranking can all be used in retrieval. The LLM will only receive passages that satisfy a predetermined relevance threshold. Instead of producing unsupported guidance, the system will declare that the controlled knowledge base does not support a trustworthy response if sufficient evidence cannot be retrieved.

The system will use RAG to combine a pretrained language model with evidence retrieved from the controlled corpus (Lewis et al., 2020). The final prompt will contain the user role, original question, retrieved passages, source metadata, response-format requirements, and safety instructions. The model will be instructed to ground substantive claims in the supplied evidence, distinguish established evidence from cautious interpretation, and communicate uncertainty when evidence is limited.

Although RAG will make source grounding and evidence updating easier, it won't be regarded as an adequate defense against hallucinations, incorrect synthesis, or retrieval errors. As a result, retrieval and generation performance will be evaluated independently and as a complete system.

The chatbot's identity, purpose, permitted and prohibited functions, evidence requirements, tone, ambiguous behavior, and safety obligations will all be specified in a global system prompt. The pathways for parents and occupational therapists will be managed using various role prompts. Answers will be given to parents using simple Greek, avoiding prescriptive or diagnostic language, and relating the information to their everyday activities and involvement. While stressing that the system does not take the place of assessment or responsible clinical judgment, professional responses may include occupational terms, evidence summaries, contextual details, and reflective thinking prompts.

Task-specific templates will be developed for concept explanations, evidence summaries, general occupational implications, professional discussion questions, and determining when a referral is required. The technical team, occupational therapists, and parent contributors will work together to develop the prompts. Version control will be used to maintain major prompt versions and revisions.

Every retrievable segment will have its own source identifier and bibliographic record. The user-facing response will show condensed source information, while an expandable panel or reference section will provide full references. The model will not be permitted to produce references from memory. The related claims will be supported by a post-generation verification process that confirms the referenced sources were actually retrieved.

At the input, generation, and output stages, safety will be enforced. Requests about diagnosis, medication, emergencies, abuse, neglect, self-harm, immediate danger, harmful practices, or replacement of professional assessment will be identified by pre-generation checks. The system may bypass normal generation and initiate a response authorized by the researcher, depending on the level of risk.

Diagnosis, medication recommendations, unsupported individual treatment plans, and unjustified certainty will all be prohibited by generation-stage instructions. Prohibited content, unsafe recommendations, diagnostic language, missing referral advice, source mismatches, and potentially identifiable information will all be evaluated by post-generation checks. Responses that don't pass checks will either be replaced with a predetermined safe response or regenerated under more stringent circumstances.

Role selection, privacy alerts, the conversational interface, source display, feedback features, and visible disclaimers about the chatbot's limitations and experimental nature will all be available on the front end. Session control, preprocessing, retrieval, prompt construction, model access, guardrails, and logging will all be handled by the back end. Following formal technical selection, the final programming languages, frameworks, libraries, hosting environment, and model-access arrangements will be reported.

Before deployment, the final programming languages, software frameworks, libraries, hosting environment, model-access route, authentication protocols, access controls, encryption arrangements, logging configuration, and security measures will be chosen. Compatibility with institutional data-protection requirements, secure processing, limited access, reproducibility, and adherence to the General Data Protection Regulation will be given top priority during the selection process. Before formal evaluation and documentation, the final technical stack and package versions will be frozen in the technical report.

A pseudonymous record comprising user role, model version, prompt version, retrieval configuration, retrieved-source identifiers, generated response, latency, safety flags, and technical failures will be produced during formal evaluation interactions. Version control will be used to manage knowledge-base versions, prompts, source documents, safety regulations, assessment scenarios, and application code. Before official evaluation, a stable release will be produced.

Reporting of model identifiers, prompt development, access route, query dates, number of independent sessions, repeated queries, reference standards, performance evaluation, and reproducibility will be guided by applicable CHART recommendations (The CHART Collaborative, 2025). Document ingestion, metadata preservation, embedding creation, retrieval relevance, prompt activation, citation accuracy, safety pathways, output stability, and adversarial requests will all be tested internally before expert evaluation. The presence of pertinent evidence in the top-ranked results will be evaluated using a manually created benchmark of Greek queries and expected source passages.

There won't be automatic parameter-level fine-tuning. RAG, prompt engineering, metadata filtering, role-specific templates, and safety regulations will all be used in the first approach. Only when stable limitations cannot be addressed by these techniques and when a sufficiently large, independently validated conversational dataset becomes available can fine-tuning be taken into consideration. Therefore, rather than training a new AI model, the first study will be defined as development, configuration, domain adaptation, knowledge grounding, and evaluation.

2.6 Phase 4: Iterative co-design and prototype refinement

The interface, role selection, response structure, source display, and safety statements will all be first shown in a low-fidelity prototype. Occupational therapists and parents will be invited to participate in separate online co-design sessions to provide feedback on navigation, terminology, response length, tone, perceived usefulness, trust, transparency, missing features, inappropriate functions, and referral communication.

Review of the low-fidelity interface and conversational flow, review of the functional minimum viable product, and final refinement before formal evaluation are the three main development cycles that are anticipated. Changes that are suggested will be noted as either accepted, accepted with modifications, deferred, or rejected, along with the justification.

2.7 Phase 5: Formative evaluation

Two different sets of scenarios will be developed. Only safety testing, retrieval optimization, prompt refinement, and internal system construction will make use of the first, which will function as a development set. The second will be a held-out evaluation set and will not be used to modify retrieval settings, safety rules, source-selection procedures, or prompts before formal evaluation.

About eight scenarios that reflect typical user queries, unclear requests, retrieval challenges, and high-risk interactions will be included in the development set. As the research team works to improve the system

architecture, these scenarios might be submitted more than once. The accuracy of the final system will not be demonstrated by performance on the development set.

The evaluation set will consist of thirty routine scenarios and ten high-risk or safety scenarios. The thirty routine scenarios will cover the two intended user roles, the three neurodevelopmental conditions, and the five main occupational and functional areas included in the minimum viable product. This distribution makes an effort to ensure that the evaluation accurately reflects the scope of the prototype. It will refrain from concentrating too much on any one occupational therapy practice area, user group, or diagnosis.

A particular coverage matrix will be used to construct the routine scenario set. The intended user role, neurodevelopmental condition, primary occupational or functional area, type of information requested, expected level of response complexity, pertinent evidence sources, and expected response behavior. The research team will review the completed matrix to find any areas that are underrepresented or duplicated prior to formal testing. If a scenario appropriately addresses multiple domains, secondary domains will be noted separately, and a primary domain will be designated for coverage.

The chatbot's ability to handle requests about diagnosis, medication advice, emergencies, safeguarding issues, potentially harmful interventions, unsupported treatment suggestions, out-of-scope questions, and inadequate contextual information will be tested in high-risk scenarios. To make sure that the held-out set accurately reflects the intended user roles, neurodevelopmental conditions, occupational domains, question types, and risk levels, a scenario-coverage matrix will be created. For every scenario, the matrix will document the intended user role, principal topic, condition, occupational domain, degree of complexity, risk category, expected retrieval sources, and expected response behavior.

No identifiable children, families, clinical records, or prior participant interactions will be used in any of the scenarios; instead, all information will be fictitious. Before the chatbot version is frozen for formal testing, members of the technical team in charge of prompt engineering and system configuration will not have access to the entire held-out evaluation set.

Instead of being statistically representative of all potential interactions with the chatbot, the scenario set will be purposeful. Examining the minimum viable product's performance across predetermined representative domains and safety categories will be its goal. As a result, the accuracy and safety results will not be extrapolated to all possible neurodevelopmental, occupational therapy, or clinical questions; rather, they will be interpreted as formative evidence about the tested scenarios. A predefined reference standard will be developed for every held-out evaluation scenario before the chatbot generates any response for formal assessment. The reference standard will specify the evidence and response characteristics against which the chatbot output will be judged.

For each scenario, the reference-standard document will contain the intended user role, the principal question or information need, the scientific sources considered directly relevant, the essential content elements expected in an acceptable response, important contextual qualifications, appropriate statements of uncertainty, and any required professional-referral or safety message.

The document will also identify statements or behaviours that would render a response unacceptable. These may include factual inaccuracies, unsupported claims, fabricated sources, inappropriate diagnostic conclusions, medication recommendations, unsafe intervention advice, excessive certainty, failure to recognise insufficient information, or omission of an essential referral or safeguarding response.

Reference standards for clinical and occupational therapy scenarios will be prepared by at least two researchers or professionals with relevant occupational therapy or neurodevelopmental expertise. Safety-critical scenarios will additionally be reviewed by a professional with appropriate expertise in clinical safety, ethics, safeguarding, digital health, or health-data governance, depending on the nature of the scenario.

Reference standards will be prepared using the same curated evidence hierarchy used for construction of the chatbot's knowledge base. However, the reference-standard team will review the complete relevant evidence rather than relying solely on passages retrieved automatically by the chatbot. This will allow the evaluation to distinguish between retrieval failure and generation failure.

Differences between reference-standard developers will be resolved through discussion. When consensus cannot be achieved, a third independent expert will review the evidence and adjudicate. The final reference standard will be approved and dated before formal chatbot outputs are generated.

The reference standards will not prescribe one mandatory sentence-by-sentence answer. Instead, they will define the minimum essential information, acceptable alternative formulations, required limitations, and prohibited content. This approach will allow natural variation in chatbot responses while maintaining a transparent basis for

evaluating factual and clinical adequacy (Figure 3).

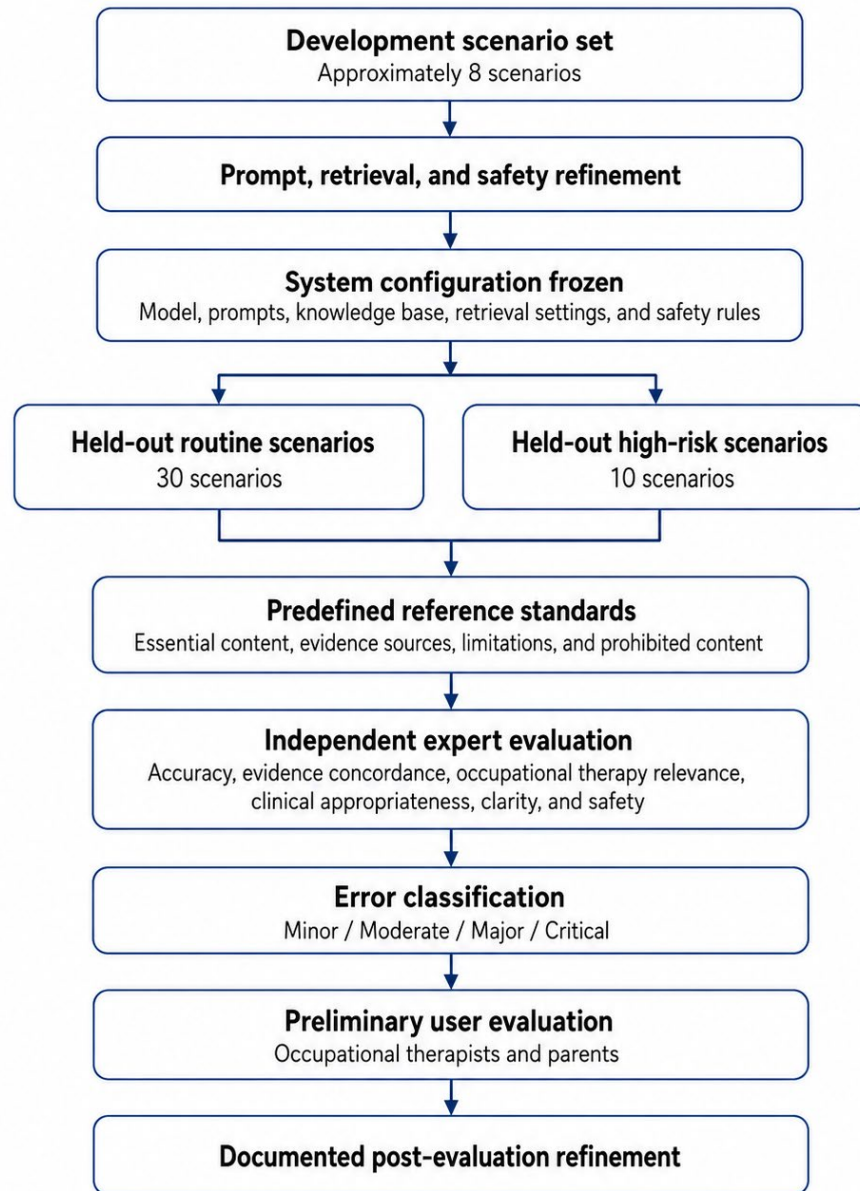


Figure 3. Planned formative evaluation pathway

Only after the knowledge base, prompt architecture, retrieval configuration, model version, safety regulations, and application release have been frozen will formal testing start. Any further changes will result in a new version of the system and necessitate retesting the impacted scenarios.

Each held-out scenario will be presented in a new conversational session to prevent information from previous exchanges influencing the response. The system will be reset in between scenarios, and no additional explanation will be provided unless the scenario explicitly evaluates the chatbot's ability to gather pertinent contextual data.

The researcher will present the scenarios in their original form, without any manual revisions. The testing date, model version, prompt version, retrieval setting, retrieved-source identifiers, response latency, final response, and activated safety rules will all be automatically recorded. A predetermined subset of routine and high-risk scenarios will be repeatedly submitted in separate sessions under the same technical conditions in order to

evaluate response stability. The final results paper will include the number of repetitions, which will be decided upon before formal evaluation.

Before expert evaluation, the research team will not manually edit, shorten, correct, or enhance the chatbot outputs. Citation verification, safety filtering, and automatic regeneration are examples of system-level post-generation processes that will be handled as a component of the chatbot system and recorded in the technical log.

The chatbot responses will be independently evaluated by an interdisciplinary panel of experts. The experts will be given roles that correspond to the dimensions being assessed. Neurodevelopmental specialists and occupational therapists will assess important components such as factual accuracy, relevance to occupational therapy, clinical appropriateness, contextual adequacy, completeness, and suitability for the intended user. Experts in AI and digital health will assess system limitations, technical coherence, retrieval behavior, and source grounding. Experts with the required background will assess the data-governance, safety, ethics, and safeguarding aspects.

The user scenario, the chatbot's response, the intended user role, and the authorized reference standard will all be sent to experts. They won't be informed about the prompt's construction, the development decisions that came before the response, or whether the response was produced before or after a specific cycle of refinement.

Five-point ratings for factual accuracy, evidence concordance, occupational therapy relevance, clinical appropriateness, clarity, completeness, role appropriateness, and safety will be applied to each response. Additionally, experts will note any unsupported claims, factual errors, clinically significant omissions, inappropriate levels of certainty, irrelevant information, misleading citations, fabricated sources, unsafe recommendations, or failure to initiate an expected refusal or referral pathway.

When a response receives a median rating of at least 4 for factual accuracy, occupational therapy relevance, clinical appropriateness, clarity, role appropriateness, and safety, and when at least 80% of pertinent expert ratings are 4 or 5, it will be deemed provisionally acceptable. Only professionals who are qualified to evaluate the relevant domain will have their ratings evaluated.

Even if a response's median score satisfies the general acceptability threshold, it will not be deemed acceptable if it contains a critical safety error.

Issues will be categorized based on how serious they are. A limited wording, completeness, or presentation issue that is unlikely to change the user's comprehension or behavior is referred to as a minor error. Incomplete, unclear, or inadequately contextualized information could be considered a moderate error. Issues will be categorized based on how serious they are. A limited wording, completeness, or presentation issue that is unlikely to change the user's comprehension or behavior is referred to as a minor error. Incomplete, unclear, or inadequately contextualized information could be considered a moderate error that could affect interpretation but is unlikely to produce immediate harm.

A major error will involve materially incorrect information, misleading evidence interpretation, inappropriate certainty, or advice that could influence professional or parental decision-making in an unsuitable manner. A critical safety error will involve content that could plausibly contribute to harm, delay urgent professional assistance, encourage unsafe intervention, provide diagnostic or medication advice outside the system's scope, fail to respond appropriately to immediate risk, or falsely present fabricated evidence as authoritative.

All major and critical errors will require mandatory review and system revision before progression to the user-testing phase. Critical errors will be reviewed individually by the research team and the relevant safety expert. Descriptive results will include medians, interquartile ranges, the proportion of ratings at or above the predefined acceptability threshold, and the frequency and severity of identified errors.

Inter-rater reliability will be examined separately for dimensions rated by the same expert group. An intraclass correlation coefficient may be used for continuous or ordinal scale ratings where the same appropriately qualified experts rate all relevant responses. Percentage agreement or an appropriate categorical agreement statistic will be used for the presence or absence of specific error types and safety failures.

Substantial disagreement among experts will not be concealed by reporting only an overall median. Responses with widely dispersed ratings will undergo structured review to determine whether disagreement reflects ambiguity in the scenario, limitations in the reference standard, disciplinary differences, or inconsistency in the chatbot response.

The development log will record any changes made following expert review. In the formal evaluation dataset,

the updated response will not take the place of the initial outcome. The new response will be reported as part of a different post-refinement evaluation if the system is retested after revision.

A 30- to 40-minute online user-testing session will be conducted with parents and occupational therapists. Before determining their intended user role and finishing four to five predetermined tasks, they will receive standardized instructions. Additionally, participants will be allowed to submit one self-generated question as long as it doesn't include any personally identifiable or health-related information.

After using the system, participants will fill out a brief study-specific acceptability questionnaire and the Greek version of the System Usability Scale. Additionally, they will respond to two open-ended questions about the system's most beneficial features and those that need to be improved or modified.

Ten items make up the System Usability Scale, which yields a total score between 0 and 100. It has been used extensively as a quick, standardized usability test, even in Greek-language studies (Lewis, 2018; Orfanou et al., 2015). Perceived utility, clarity, trust, perceived safety, relevance, and intention to use will all be evaluated by the study-specific questionnaire. There will be no need for a follow-up period, repeated clinical evaluation, or longitudinal data collection.

2.8 Study outcomes

The completion of a stakeholder-informed minimum viable product, the development of a requirements-traceability matrix, the creation of distinct conversational pathways for parents and occupational therapists, the construction of a curated knowledge architecture, the implementation of predetermined safety mechanisms, and the documentation of the modifications made throughout prototype iterations are the main development outcomes. The primary evaluation outcomes will be expert-rated factual accuracy, occupational therapy relevance, clinical appropriateness, management of high-risk scenarios, and user-rated usability as measured by the System Usability Scale.

Secondary outcomes will include evidence concordance, clarity, completeness, role appropriateness, frequency of hallucinations or unsupported claims, retrieval and source-attribution errors, task-completion rates, perceived usefulness, acceptability, trust, perceived safety, intention to use, and qualitative recommendations for improvement (Table 2).

Table 2. Planned study outcomes.

Outcome level	Principal domains	Planned indicators
Development	Traceability and completion	Minimum viable product, requirements matrix, role pathways, knowledge architecture, safety rules, documented revisions
Expert	Quality and professional relevance	Accuracy, evidence concordance, occupational therapy relevance, clinical appropriateness, clarity, completeness, role appropriateness
Technical	Retrieval and generation performance	Retrieval relevance, source-to-claim consistency, citation accuracy, response failures, latency, stability
Safety	Risk recognition and response	Unsafe-response rate, correct refusal or referral, prohibited diagnostic or medication language, activation of safety pathway
User	Formative user experience	SUS score, task completion, usefulness, acceptability, trust, perceived safety, intention to use, qualitative feedback

Source: Authors, 2026.

2.9 Data analysis

Focus-group, co-design, and open-ended feedback data will be analysed using the Framework Method, which is appropriate for multidisciplinary health research involving professional and lay stakeholders (Gale et al., 2013). The process will include familiarisation, coding, development and application of an analytical framework, charting into a matrix, and interpretation within and across stakeholder groups. At least two researchers will contribute to framework development and review.

Quantitative analyses will primarily be descriptive because the study is formative and involves relatively small samples. Frequencies and percentages will be reported for categorical variables, while means and standard deviations or medians and interquartile ranges will be reported for continuous and ordinal variables, as appropriate. Task completion rates, the percentage of responses that satisfy the predetermined acceptability criteria, System Usability Scale scores, retrieval and citation errors, and the frequency of minor, moderate, major, and critical errors will all be analyzed.

The median, interquartile range, complete observed range, and percentage of ratings of 4 or 5 will be reported for each expert-rated dimension. When a chatbot response has a median rating of at least 4, and at least 80% of the relevant expert ratings are 4 or 5, it will be deemed provisionally acceptable. Only experts whose professional or technical expertise is suitable for the dimension being evaluated will have their ratings included.

Only when all pertinent responses are rated by the same set of suitably qualified experts will inter-rater reliability be investigated independently for each evaluation domain. When scale-based ratings are supported by the rating design's assumptions and structure, intraclass correlation coefficients will be computed. In compliance with the recommendations of Koo and Li (2016), the ICC model, type, and definition will be chosen based on whether the raters are treated as a fixed or random sample and whether reliability is based on single-rater or average-rater measurements. These choices will be clearly reported.

When possible, percentage agreement and an appropriate chance-corrected agreement statistic will be reported for categorical outcomes, such as the existence or absence of unsupported claims, factual errors, unsafe recommendations, citation failures, and required referral or refusal behaviors. Clinical interpretation will not be replaced by agreement statistics, and responses exhibiting significant expert disagreement will be subjected to a systematic review.

There are no planned confirmatory hypothesis tests or formal statistical comparisons between parents and occupational therapists. User group differences will be explained and interpreted cautiously. A joint-display matrix connecting numerical performance, stakeholder explanations, identified design issues, suggested changes, and final decisions will be used to integrate quantitative and qualitative findings. This will give an auditable record of how user input and evidence shaped the finished prototype

2.10 Data management, ethics, and public involvement

Regulation (EU) 2016/679, the General Data Protection Regulation, and the relevant institutional research-governance and data-protection requirements will all be followed by this experimental proof-of-concept study. Names, dates of birth, addresses, schools, record numbers, photos, clinical reports, and any other identifying information will not be allowed to be entered by participants. Pseudonymized research data will be kept on password-protected, institutionally approved systems. Research data and consent documentation will be kept apart.

Before hiring, ethical approval will be secured. In addition to providing electronic informed consent, participants will be informed that the system is experimental, does not diagnose, does not replace assessment, and does not provide a direct clinical benefit. Unsafe answers found during testing will be recorded, examined, and fixed before additional assessment.

The study protocol will be submitted for ethical review to the Committee of Ethics and Deontology in Research of the University of the Aegean. Participant recruitment and data collection will not commence until written ethical approval has been obtained. Participant recruitment and data collection will not begin before written ethical approval has been obtained.

Parents will contribute as co-design partners as well as research participants. Their input will influence language, content priorities, privacy messages, interaction pathways, boundaries, and safety communication. The final report will identify system features that were directly informed by parent and occupational therapist contributions.

2.11 Reporting and study status

The final development and evaluation report will apply relevant CHART items for transparent documentation of model selection, prompts, query procedures, evaluation standards, reproducibility, and data availability (The CHART Collaborative, 2025). Deviations from the protocol and changes to the model, knowledge base, or safety architecture will be documented. At the time of manuscript submission, the study is at the protocol-development

and technical-planning stage. Participant recruitment and formal data collection have not yet commenced and will begin only after ethical approval has been obtained.

3. Anticipated Outputs and Results

The study is expected to aid in the methodical creation of a conversational artificial intelligence prototype that is evidence-based, stakeholder-informed, and created especially for parents of children with neurodevelopmental disorders and occupational therapists. The goal of the study is to establish the methodological, technical, and evaluative foundations for subsequent empirical testing rather than to demonstrate clinical efficacy.

We anticipate that the phases of contextual inquiry and stakeholder consultation will identify common as well as role-specific needs among parents and occupational therapists. Occupational therapists are expected to demand more thorough evidence summaries, professional terminology, and context-sensitive clinical-reasoning support, while parents are expected to prioritize understandable explanations, practical relevance, and protection from inappropriate or dangerous guidance. These anticipated differences will inform the development of distinct conversational pathways for the two user groups.

The evidence-synthesis process is expected to produce a structured knowledge architecture covering the principal domains included within the initial scope of the system. Executive functioning, self-regulation, sensory and emotional regulation, motor coordination, self-care, play, education, social engagement, family routines, environmental adaptation, and family-centered occupational therapy are some of these domains. The initial focus of condition-specific content will be on developmental coordination disorder, attention-deficit/hyperactivity disorder, and autism spectrum disorder.

The creation of a requirements-traceability matrix that connects particular technical, conversational, clinical, and safety features with stakeholder priorities and scientific evidence will be a key product of the study. It is anticipated that this matrix will offer an open record of the identification, justification, implementation, and evaluation of each major system function. It will also list requirements that were postponed or dropped because they went beyond the prototype's intended scope or raised issues with safety, feasibility, ethics, or evidence.

It is anticipated that the iterative co-design process will result in a minimum viable product in Greek with distinct interaction pathways for parents and occupational therapists. A curated evidence base, retrieval-augmented generation, structured system instructions, user-role adaptation, source-grounded responses, explicit statements of uncertainty, and predetermined safety and referral mechanisms will all be features of the prototype.

It is anticipated that the system will assist users in obtaining and analyzing evidence-based data regarding neurodevelopmental functioning and occupational participation. It won't, however, diagnose conditions, recommend medication adjustments, handle emergencies, or take the place of expert evaluation and clinical judgment. Requests that go beyond its intended scope will be identified by the prototype, which will then offer a suitable limitation statement, refusal, request for non-identifiable clarification, or referral recommendation.

Preliminary evidence regarding the generated responses' factual accuracy, evidence concordance, occupational therapy relevance, clinical appropriateness, clarity, completeness, role appropriateness, and safety is anticipated to be produced during the expert-evaluation phase. Additionally, unsupported claims, factual errors, inappropriate certainty, clinically significant omissions, source-attribution errors, and failures to activate predefined safety pathways are expected to be identified by the evaluation. Before a more comprehensive assessment, the system will be formatively revised using these findings.

Technical testing should yield data on retrieval performance, latency, source-attribution accuracy, response variability, consistency between generated responses and retrieved evidence, and the dependability of safety mechanisms. An early indication of the stability and repeatability of the model's responses under standardized conditions will also be obtained through repeated testing of specific scenarios.

The system's usability, acceptability, perceived usefulness, clarity, trustworthiness, perceived safety, and relevance to the needs of occupational therapists and parents are all expected to be described in preliminary user evaluations. It is anticipated that qualitative feedback will point out particular aspects that need improvement, especially with regard to response length, terminology, evidence presentation, navigation, transparency, and the disclosure of system limitations.

A final development matrix that links quantitative evaluation results with qualitative explanations and the corresponding design modifications will be produced as a result of the mixed-methods integration process. This

procedure will give a clear explanation of how input from experts and stakeholders affected the final prototype.

A documented requirements framework, role-specific conversational pathways, a functional Greek-language prototype, a safety and escalation architecture, a standardized set of routine and high-risk evaluation scenarios, a preliminary evaluation protocol, and a curated and structured knowledge base will thus be the main results of the study.

Whether the prototype has attained a level of functionality, occupational therapy relevance, safety, usability, and acceptability sufficient to support advancement to larger feasibility and effectiveness studies will be determined by the results of the planned expert, technical, safety, and user evaluations. A separate results paper will present the empirical findings from these formative assessments.

4. Discussion

In order to support occupational therapists and parents of children with neurodevelopmental disorders, the proposed study will create an organized methodology for the co-design and development of an evidence-based conversational AI system. Its main contribution will be the clear conversion of scientific evidence, stakeholder priorities, occupational therapy principles, and safety requirements into a working digital prototype rather than the demonstration of clinical efficacy.

The study fills in several gaps found in the body of current literature. Although the evidence that is currently available is still inconsistent and often comes from small-scale or early-stage studies, conversational AI has demonstrated potential to enhance access to health information, support education, and enable interactive guidance (Busch et al., 2025; Milne-Ives et al., 2020). Existing conversational systems in neurodevelopmental care have primarily concentrated on information access, engagement, parental support, or condition-specific assistance rather than professional reasoning, family routines, occupational performance, or participation (Bui et al., 2022; Singla and associates, Zhang et al., 2024, 2026). By combining an occupational therapy-focused framework with distinct conversational pathways for professionals and parents, the suggested system aims to expand this body of literature. This dual-user design acknowledges that while the two groups may pose similar queries, their needs for terminology, specific evidence, contextual interpretation, and helpful advice may differ. While occupational therapists might need more thorough evidence summaries, professional terminology, and prompts that support rather than replace clinical reasoning, parents might benefit from clear explanations related to daily routines and participation.

Because general-purpose language models cannot be expected to replicate the contextual, occupation-focused reasoning of occupational therapists, this distinction is especially crucial. Değerli and associates (2026) discovered that while ChatGPT and occupational therapists showed agreement in certain areas of assessment and intervention selection, there were notable differences in other areas. These results imply that the interaction between the child, the environment, work demands, family priorities, and the therapist's professional judgment may not be adequately taken into account by unrestricted chatbot responses. Therefore, rather than being an independent clinical decision-maker, the suggested system will be built as a source-grounded educational and decision-support tool.

4.1 Contribution to neurodevelopmental and cognitive support

The system's conceptual architecture will also build upon emerging evidence concerning the relationship between conversational AI and cognitive or metacognitive functioning. Pergantis et al. (2025a) identified potential applications of AI chatbots in areas such as attention, working memory, cognitive flexibility, problem-solving, self-regulation, decision-making, and metacognition. Although the available evidence does not establish therapeutic effectiveness, it supports further investigation of conversational processes involving structured feedback, reflection, planning, and adaptive guidance.

The proposed chatbot will not be presented as an intervention that directly improves executive functions. Instead, evidence concerning cognitive and metacognitive support will inform the organisation of information, the structure of reflective prompts, and the development of guidance relating to planning, routines, problem-solving, and self-regulation. Any future claims concerning child-level cognitive improvement would require separate prospective evaluation using validated outcome measures.

The inclusion of sensory and emotional regulation content is similarly supported by previous research examining emerging technologies in autism and sensory-processing difficulties. Pergantis et al. (2025b) identified growing

use of AI-supported monitoring, wearable devices, virtual environments, biofeedback, mobile technologies, and other assistive systems in the detection or management of stress and anxiety. These findings support the broader relevance of technology-assisted regulation strategies but do not demonstrate that a conversational system can independently assess or treat neurophysiological stress. Accordingly, the proposed chatbot will limit its role to evidence-informed education, environmental considerations, routine-based guidance, and appropriate referral recommendations.

The study also responds to the assessment-oriented emphasis identified within AI research on Developmental Coordination Disorder. Pergantis et al. (2026) discovered that, with comparatively little focus on intervention, participation, and rehabilitation support, AI applications in this field have primarily addressed screening, classification, movement analysis, and assessment. By concentrating on the interpretation and dissemination of data pertinent to participation, daily activities, family routines, and occupational performance, the suggested system will investigate an alternative translational path.

4.2 Methodological contribution

The protocol's integration of frameworks for digital health, complex intervention, technological development, and participatory design is a key methodological strength. The main framework for contextual inquiry, value specification, design, operationalization, and evaluation will be provided by the CeHRes Roadmap 2.0. The system will be positioned as a complex intervention that needs to go through development and feasibility stages before its efficacy can be examined, according to the Medical Research Council framework. While co-design and user-centered approaches will guarantee that the system is shaped by the priorities of its intended users, Design Science Research will direct the creation and assessment of the digital artifact.

The risk of a technology-driven development process, where system functions are chosen mainly based on technical capability, is anticipated to be decreased by this integration. Rather, every significant feature will be connected to evidence, a need identified by stakeholders, a professional requirement, or a safety consideration. The development log and requirements-traceability matrix will offer an open documentation of this procedure.

Additionally, the study makes use of indexed methodological precedents in the creation of chatbots. Livermore and associates (2024) outlined a co-designed proof-of-concept protocol that involved assessing stakeholder needs, creating a minimum viable product, and evaluating users iteratively before moving on to a more extensive efficacy study. The current protocol extends this staged logic to neurodevelopmental care and occupational therapy.

The decision to develop the system initially in Greek represents both a contextual contribution and a methodological challenge. Most large language models and digital-health systems are developed and tested primarily in English. A Greek-language occupational therapy chatbot may therefore improve accessibility within the national context, but its linguistic accuracy, interpretation of professional terminology, and ability to communicate complex evidence in accessible Greek will require careful evaluation.

4.3 Safety, transparency, and human oversight

One of the biggest obstacles to the development of the suggested system is anticipated to be safety. Users may view LLM-generated responses as more trustworthy or authoritative than they actually are due to their conversational fluency. Concerns about fabricated information, bias, privacy, lack of transparency, unclear accountability, and the creation of believable but dangerous advice have been noted in previous reviews (Busch et al., 2025; Ranisch; Haltaufderheide, 2024).

A combination of source-grounded generation, predetermined system boundaries, role-specific communication, expert review, high-risk scenario testing, and explicit declarations that the system does not replace professional assessment will be used by the proposed system to mitigate these risks. Although responses will be linked to a carefully selected body of evidence using retrieval-augmented generation, it will not be regarded as a comprehensive remedy for source errors or hallucinations. Although the quality of retrieval, source interpretation, and final response generation must still be independently assessed, prior research has indicated that retrieval-augmented generation may strengthen the relationship between generated content and domain-specific evidence (Miao et al., 2024).

Human oversight will remain essential throughout development and subsequent use. Occupational therapists will retain responsibility for interpreting evidence within the context of individual children and families. Parents will

be advised that the system offers general educational guidance and cannot determine whether a particular recommendation is appropriate for their child. Requests involving diagnosis, medication, emergencies, immediate risk, or potentially harmful interventions will activate limitation, refusal, or referral pathways.

4.4 Strengths of the proposed study

The planned study is expected to have several advantages. First, instead of depending only on general model knowledge or unrestricted internet content, it will be supported by current evidence. Clinical guidelines, systematic reviews, scoping reviews, occupational therapy frameworks, and pertinent indexed primary studies will all be incorporated into the knowledge architecture.

Second, parents and occupational therapists will be involved at every stage of the development process. It is anticipated that their participation will enhance the system's contextual appropriateness, acceptability, accessibility, and relevance. This is in line with earlier studies demonstrating that intended users may have varying preferences for chatbot features, personalization, information limitations, accuracy, and trust (Bui et al., 2022).

Third, professional and parent-facing communication will be clearly distinguished in the study. The system will try to account for variations in role, responsibility, terminology, and informational needs rather than just simplifying the same response for parents.

Fourth, rather than treating safety evaluation as an activity added after prototype completion, the study will incorporate safety by design. During the requirements and architecture phases, specifications will be made regarding system boundaries, high-risk requests, refusal mechanisms, uncertainty communication, and professional referral pathways. Lastly, the study will record how evidence, stakeholder needs, system requirements, prototype features, and evaluation results relate to each other. This traceability could improve the development process's reproducibility and offer a methodological model for upcoming AI research in occupational therapy and rehabilitation.

4.5 Anticipated limitations

There are a few restrictions to take into account. Purposive and convenience sampling will be the main methods used to recruit stakeholders while the system is first developed and assessed in Greece. As a result, participants might be more interested in digital technologies or research than parents and occupational therapists in general. Not all intended users may be represented in the requirements and usability findings that follow.

Because the goal of the study is formative development rather than population-level inference, relatively small samples will be used. As a result, the study will not be able to determine statistically significant differences between occupational therapists and parents, and quantitative results will be interpreted descriptively.

Three neurodevelopmental disorders and five predetermined occupational and functional domains will be addressed by the first minimum viable product. The full range of neurodevelopmental conditions, comorbidities, occupational needs, and clinical questions encountered in pediatric occupational therapy will not be represented by this limited scope, despite the fact that it is meant to increase methodological feasibility and the depth of evidence curation. Although the formal evaluation scenarios will be purposefully created to reflect the specified prototype scope, they will not offer a comprehensive evaluation of all potential user interactions. Therefore, results will only be interpreted in light of the domains, user roles, safety categories, and tested conditions.

The quality of the system will also depend on the performance and stability of the underlying language model. Changes made by the model provider, including updates to model architecture, safety rules, or data-processing procedures, may affect reproducibility. The model version and evaluation dates will therefore be documented, and major changes during the formal evaluation period will be avoided where possible.

The evidence base may become outdated as new research and clinical guidance emerge. A structured updating procedure will therefore be necessary. However, the initial protocol will not establish the long-term resources, governance structures, and responsibilities required for continuous maintenance after completion of the research study.

Additionally, ambiguity, missing context, or terminology that deviates from the language used in the knowledge base may be present in user questions. Therefore, even in cases where the included sources are scientifically valid, the system may retrieve information that is either incomplete or irrelevant. Although source-to-response concordance and retrieval quality will be assessed technically, these procedures are unable to completely

eradicate errors.

Neither clinical decision-making in the real world nor access to clinical records will be used to assess the system with children. As a result, the results will not show clinical benefit, efficacy, or readiness for implementation. Only whether the prototype is accurate, pertinent, safe, usable, and acceptable enough to support moving forward with further research will be indicated.

4.6 Implications for the Greek occupational therapy context

The planned study might be especially important for the advancement of digital innovation and occupational therapy research in Greece. Onoufriou et al. (2027) discovered that while Greek occupational therapy research was scarce for several decades, it showed significant expansion after 2021. Emerging technologies, neurodevelopment, cognitive functioning, and rehabilitation were identified among the developing thematic areas.

The present study may contribute to this emerging direction by developing a Greek-language digital tool grounded specifically in occupational therapy. It may also strengthen collaboration among occupational therapists, parents, researchers, and technology specialists. However, its national relevance should not be interpreted as sufficient evidence of international novelty. The contribution of the study must remain grounded in comparison with internationally indexed research concerning conversational AI, occupational therapy, digital rehabilitation, and neurodevelopmental care.

4.7 Future research

The results of the expert, technical, safety, usability, and acceptability evaluations outlined in the current protocol will be reported in the very next publication resulting from this research program. These formative assessments will be carried out in Greece with a small panel of experts, small groups of parents and occupational therapists, minimal participant burden, and online participation.

Clinical efficacy and child-level results will not be examined in the formative assessment. The prototype must meet predetermined criteria for factual accuracy, evidence concordance, occupational therapy relevance, clinical appropriateness, safety, usability, and acceptability to advance to larger feasibility or effectiveness studies.

The occupational therapy-focused system may be compared to a general-purpose conversational AI system or traditional online information sources in later studies. Factual accuracy, source grounding, occupational therapy relevance, safety, user trust, task completion, and the effectiveness of evidence retrieval are examples of relevant outcomes. Prospective studies should only look at parental knowledge, professional confidence, occupational performance, participation, executive functioning, family routines, parental stress, or other clinical and functional outcomes after sufficient technical performance, safety, and feasibility have been established.

5. Conclusions

This protocol describes the planned co-design, technical development, and preliminary evaluation of an evidence-based occupational therapy-focused conversational AI system for occupational therapists and parents of children with neurodevelopmental disorders. The study will combine curated scientific evidence, RAG, role-specific prompt engineering, source verification, layered safety guardrails, and stakeholder-informed design.

The proposed system will be an educational and decision-support resource, not an autonomous diagnostic or therapeutic agent. Transparent documentation of the evidence base, technical architecture, prompt structure, model configuration, safety procedures, and evaluation methods will support the planned formative evaluation in Greece. The findings of this evaluation will be reported in a separate results paper and will determine whether progression to larger feasibility and effectiveness studies is justified.

6. Authors' Contributions

Conceptualization: P.P.; Methodology: P.P., N.B., C.S.; Study design: P.P.; Evidence synthesis: P.P.; Technical and evaluation framework: P.P.; Visualization: P.P.; Project administration: P.P.; Supervision: C.S., A.D.; Validation: P.P., N.B., A.D.; Writing—original draft: P.P.; Writing—review and editing: P.P., N.B., C.S., A.D. All authors have read and approved the final version of the manuscript.

7. Conflicts of Interest

No conflicts of interest.

8. Ethics Approval

Ethical approval is pending. The study will not commence until written approval has been obtained from the Committee of Ethics and Deontology in Research of the University of the Aegean.

9. Artificial Intelligence Statement: Declaration of Generative AI and AI-Assisted Technologies in the Writing Process

Not applicable. The authors used OpenAI ChatGPT during the preparation of this manuscript solely to support language editing and translation to Portuguese and improve clarity and academic expression.

Appendix A. Technical Specifications to Be Finalised Before Formal Evaluation

- ☐ Base model provider, exact model and release/version
- ☐ Model access route and contractual data-retention conditions
- ☐ Generation parameters, including temperature and maximum output length
- ☐ Embedding model and cross-lingual benchmark results
- ☐ Document chunking strategy, size, and overlap
- ☐ Vector database or similarity index
- ☐ Similarity metric, top-k retrieval value, relevance threshold, and reranking method
- ☐ Programming languages, frameworks, libraries, and package versions
- ☐ Hosting environment, security controls, and institutional data-processing arrangements
- ☐ Full system prompt and role-specific prompt templates
- ☐ High-risk classification rules and researcher-approved safety responses
- ☐ Knowledge-base update and governance procedure
- ☐ Technical test set and expected source passages
- ☐ Frozen application release and evaluation dates

10. References

- Bui, T. A., Pohl, M., Rosenfelt, C., Ogourtsova, T., Yousef, M., Whitlock, K., Majnemer, A., Nicholas, D., Demmans Epp, C., Zaiane, O., & Bolduc, F. V. (2022). Identifying potential gamification elements for a new chatbot for families with neurodevelopmental disorders: User-centered design approach. *JMIR Human Factors*, 9(3), e31991. <https://doi.org/10.2196/31991>
- Bulan, P. M. P., Kuizon, D. A. Y., Casaña, R. S. E., Fuentes, C. G., Pestaño, N. Y., & Suerte, J. R. O. (2025). A scoping review on artificial intelligence in occupational therapy. OTJR: Occupation, Participation and Health. *Advance Online Publication*. <https://doi.org/10.1177/15394492251379332>
- Busch, F., Hoffmann, L., Rueger, C., van Dijk, E. H. C., Kader, R., Ortiz-Prado, E., Makowski, M. R., Saba, L., Hadamitzky, M., Kather, J. N., Truhn, D., Cuocolo, R., Adams, L. C., & Bressemer, K. K. (2025). Current applications and challenges in large language models for patient care: A systematic review. *Communications Medicine*, 5. <https://doi.org/10.1038/s43856-024-00717-2>
- Değerli, Y., Özata Değerli, M. N., & Kars, S. (2026). Artificial intelligence in occupational therapy: Comparing ChatGPT and occupational therapist approaches in assessment and intervention. *British Journal of Occupational Therapy*, 89(1), 6-18. <https://doi.org/10.1177/03080226251368231>
- Feng, F., Yang, Y., Cer, D., Arivazhagan, N., & Wang, W. (2022). Language-agnostic BERT sentence embedding. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: Volume 1, Long Papers, 878-891 p. *Association for Computational Linguistics*.

<https://doi.org/10.18653/v1/2022.acl-long.62>

- Gale, N. K., Heath, G., Cameron, E., Rashid, S., & Redwood, S. (2013). Using the Framework Method for the analysis of qualitative data in multi-disciplinary health research. *BMC Medical Research Methodology*, 13. <https://doi.org/10.1186/1471-2288-13-117>
- Haltaufderheide, J., & Ranisch, R. (2024). The ethics of ChatGPT in medicine and healthcare: A systematic review on large language models. *NPJ Digital Medicine*, 7. <https://doi.org/10.1038/s41746-024-01157-x>
- Kaelin, V. C., Valizadeh, M., Salgado, Z., Parde, N., Khetani, M. A., & Foster, E. R. (2021). Artificial intelligence in rehabilitation targeting the participation of children and youth with disabilities: Scoping review. *Journal of Medical Internet Research*, 23(11), e25745. <https://doi.org/10.2196/25745>
- Kip, H., Beerlage-de Jong, N., van Gemert-Pijnen, L. J. E. W. C., & Kelders, S. M. (2025). The CeHRes Roadmap 2.0: Update of a holistic framework for development, implementation, and evaluation of eHealth technologies. *Journal of Medical Internet Research*, 27, e59601. <https://doi.org/10.2196/59601>
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155-163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Lewis, J. R. (2018). The System Usability Scale: Past, present, and future. *International Journal of Human-Computer Interaction*, 34(7), 577-590. <https://doi.org/10.1080/10447318.2018.1455307>
- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-T., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems*, 33, 9459-9474.
- Livermore, P., Kupiec, K., Wedderburn, L. R., Knight, A., Solebo, A. L., Shafran, R., Robert, G., Sebire, N. J., Gibson, F., & IMPACT Steering Group. (2024). Designing, developing, and testing a chatbot for parents and caregivers of children and young people with rheumatological conditions (the IMPACT study): Protocol for a co-designed proof-of-concept study. *JMIR Research Protocols*, 13, e57238. <https://doi.org/10.2196/57238>
- Milne-Ives, M., de Cock, C., Lim, E., Harper Shehadeh, M., de Pennington, N., Mole, G., Normando, E., & Meinert, E. (2020). The effectiveness of artificial intelligence conversational agents in health care: Systematic review. *Journal of Medical Internet Research*, 22(10), e20346. <https://doi.org/10.2196/20346>
- Onoufriou, A., Kargaki, I., Sfintsioti, P., Georgiou, K., Kalogirou, G., & Pergantis, P. (2027). Bibliometric and science mapping analysis of occupational therapy research in Greece (1977–2025). *Brazilian Journal of Science*, 6(1), 191-207. <https://doi.org/10.14295/bjs.v6i1.931>
- Orfanou, K., Tselios, N., & Katsanos, C. (2015). Perceived usability evaluation of learning management systems: Empirical evaluation of the System Usability Scale. *The International Review of Research in Open and Distributed Learning*, 16(2), 227-246. <https://doi.org/10.19173/irrodl.v16i2.1955>
- Peppers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, 24(3), 45-77. <https://doi.org/10.2753/MIS0742-1222240302>
- Pergantis, P., Bamicha, V., Skianis, C., & Drigas, A. (2025a). AI chatbots and cognitive control: Enhancing executive functions through chatbot interactions: A systematic review. *Brain Sciences*, 15(1). <https://doi.org/10.3390/brainsci15010047>
- Pergantis, P., Bamicha, V., Doulou, A., Christou, A. I., Bardis, N., Skianis, C., & Drigas, A. (2025b). Assistive and emerging technologies to detect and reduce neurophysiological stress and anxiety in children and adolescents with autism and sensory processing disorders: A systematic review. *Technologies*, 13(4). <https://doi.org/10.3390/technologies13040144>
- Pergantis, P., Georgiou, K., Bardis, N., Skianis, C., & Drigas, A. (2026). Artificial intelligence in the evaluation and intervention of Developmental Coordination Disorder: A scoping review of methods, clinical purposes, and future directions. *Children*, 13(2). <https://doi.org/10.3390/children13020161>
- Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, 3982-3992 p., Association for Computational Linguistics. <https://doi.org/10.18653/v1/D19-1410>

- Singla, A., Khanna, R., Kaur, M., Kelm, K., Zaiane, O., Rosenfelt, C. S., Bui, T. A., Rezaei, N., Nicholas, D., Reformat, M. Z., Majnemer, A., Ogourtsova, T., & Bolduc, F. (2024). Developing a chatbot to support individuals with neurodevelopmental disorders: Tutorial. *Journal of Medical Internet Research*, 26, e50182. <https://doi.org/10.2196/50182>
- Skivington, K., Matthews, L., Simpson, S. A., Craig, P., Baird, J., Blazeby, J. M., Boyd, K. A., Craig, N., French, D. P., McIntosh, E., Petticrew, M., Rycroft-Malone, J., White, M., & Moore, L. (2021). A new framework for developing and evaluating complex interventions: Update of Medical Research Council guidance. *BMJ*, 374. <https://doi.org/10.1136/bmj.n2061>
- The CHART Collaborative. (2025). Reporting guideline for chatbot health advice studies: The CHART statement. *JAMA Network Open*, 8(8), e2530220. <https://doi.org/10.1001/jamanetworkopen.2025.30220>
- Zhang, W., Leung, R. Y. F., Mak, K. K., Ge, H., Kwok, T. T. O., Tso, R. V. Y., Wang, A., Ma, H., & Wong, J. Y. H. (2026). GPT-powered chatbot-based positive psychology intervention for well-being among parents of children with autism spectrum disorder: Single-arm mixed methods study. *JMIR Formative Research*, 10, e85060. <https://doi.org/10.2196/85060>

Funding

Not applicable.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).