Original Paper

Psychoeducational Chatbot (Demetra) for Preventing Childhood Overweight: Development and Usability Study

Barbara Purin[1], MSc; Anna Elena Nicoletti[1], BSc, MSc; Marco Bolpagni[1,2], BSc, MSc; Simone De Carli[1], BSc; Lara Devilli[3,4,5], MSc, MD; Silvia Rizzi[1], BSc, MSc, PsyD; Stefania Poggianella[1], BSc, MSc; Chiara De Luca[6], BSc, MSc; Stefano Stagi[4,5], MSc, MD; Roberto Franceschi[3,7], MSc, MD

[1]Digital Health Research, Center for Digital Health and Wellbeing, Fondazione Bruno Kessler, Trento, Trentino-Alto Adige, Italy
[2]Department of General Psychology, University of Padua, Padova, Veneto, Italy
[3]Department of Pediatrics, Azienda sanitaria universitaria integrata del Trentino, Trento, Italy
[4]Department of Health Sciences, University of Florence, Firenze, Toscana, Italy
[5]Auxo-endocrinology Unit, Meyer Children's Hospital IRCCS, Florence, Italy
[6]Pavoniano Artigianelli Institute, Trento, Trentino-Alto Adige, Italy
[7]Department of Medicine and Health Sciences (DisMed), University of Trento, Trento, Italy

Corresponding Author:

Barbara Purin, MSc
Digital Health Research, Center for Digital Health and Wellbeing
Fondazione Bruno Kessler
Via Sommarive, 18
Trento, Trentino-Alto Adige 38123
Italy
Phone: 39 0461314011
Email: purin@fbk.eu

Abstract

Background: Childhood overweight is a major public health issue in Italy, with a high prevalence among very young children and links to long-term physical and psychological harms. Family-centered approaches such as motivational interviewing in primary care show behavioral gains but struggle with scalability, resource needs, and sustained engagement. To overcome these barriers, conversational AI chatbots offer a promising route for accessible, scalable early prevention support.

Objective: This study used the Obesity-Related Behavioral Intervention Trials (ORBIT) model (phases Ia-Ib) to co-design, refine, and evaluate Demetra, a hybrid scripted and generative AI chatbot aimed at preventing overweight in at-risk children aged 2 to 3 years. The goal was to assess the acceptability, usability, and engagement of the prototype and to identify necessary refinements to inform planning for ORBIT phase IIa. Demetra draws from an existing motivational interviewing program and incorporates Acceptance and Commitment Therapy and Brief Action Planning as its core mechanisms to promote change.

Methods: A multidisciplinary team applied a user-centered design approach to develop a prototype with 4 asynchronous sessions on values clarification, psychoeducation (nutrition, physical activity, and screen time), and goal setting. In a convergent mixed methods study, 32 Italian stakeholders (pediatricians, psychologists, AI/communication experts, and families) used the prototype in scripted scenarios, rated it on quantitative scales (user version of the Mobile Application Rating Scale [uMARS] subscales, Comprehensive Evaluation Scale for LLM-Powered Counseling Chatbots [CES-LCC], semantic differentials, and ad hoc items for educational media), and joined semistructured interviews. Quantitative outcomes were compared with neutral scale midpoints using 1-tailed Wilcoxon signed-rank tests, and qualitative data were analyzed thematically.

Results: Ratings were consistently positive and significantly above neutral. uMARS scores reflected strong information delivery, perceived impact, engagement, and subjective quality. Conversational quality (CES-LCC and semantic differentials) was strong for clarity, relevance, guidance, empathy, fluency, and session length. Nonaudio educational materials, such as PDFs and images, scored highly in terms of content value, engagement, presentation, and structure, while podcasts scored well on content and engagement but only adequately on presentation and structure. In interviews, stakeholders valued linguistic clarity, ease of use, age-appropriateness for those aged 2 to 3 years, and fit with daily routines. They suggested adding practical examples, more nuanced emotional responses, and more sustained motivational prompts. They positioned Demetra as a useful supplement to pediatric care, extending the work of pediatricians beyond clinic visits.

Conclusions: Feedback confirms Demetra's acceptability, usability, and fit with family-centered prevention in a scalable asynchronous format. Suggested refinements, such as practical scenarios, adaptive emotional support, and sustained engagement tools, lay the groundwork for optimization and ORBIT phase IIa feasibility testing in pediatric settings.

JMIR Pediatr Parent 2026;9:e95697; doi: 10.2196/95697

Keywords: pediatric overweight prevention; childhood overweight; digital therapeutics; conversational AI; user-centered design; usability; engagement; mobile health

Introduction

Childhood overweight and obesity represent pressing public health concerns, with early-onset conditions linked to significant long-term physical [1] and psychological consequences [2]. Within Italy, the prevalence rates are especially elevated [3], even among infants, and recent data from our country (the Autonomous Province of Trento) indicate that approximately 1 in 5 children aged 24 months are already above the 85th BMI percentile, placing them at an elevated risk for persistent weight-related challenges throughout life [4].

The early years of life represent a critical time for intervention, during which routines related to food, movement, and emotional regulation are rapidly established [5]. Family-centered programs delivered during this period can assist in the promotion of healthier trajectories [6]. However, traditional approaches, which frequently need in-person counseling, are resource-intensive and difficult to scale [7]. Furthermore, these measures may not always align with the daily realities and emotional readiness of families [8]. Despite their evident clinical efficacy, such interventions often encounter difficulties with long-term engagement and practical implementation in real-world settings.

A pediatrician-led motivational interviewing (MI) program [9] developed in the Reggio Emilia region of Italy showed that MI delivered by trained primary care providers improves lifestyle behaviors among children aged 4 to 7 years, but only modestly reduces BMI gain [10]. As is the case with many in-person models, its effectiveness was influenced by sociodemographic factors and required sustained provider involvement, which indeed limited its scalability.

Consistent with the observations in Reggio Emilia, the broader literature on MI in pediatric care indicates primarily behavioral improvements and generally small or mixed effects on BMI, as has been confirmed in a more recent study [11] and by a cluster-randomized trial [12] conducted in European primary care settings, which revealed favorable changes in diet and lifestyle, yet no clinically significant differences in BMI. Similarly, the High Five for Kids study [13], a randomized controlled trial conducted in US primary care, reduced screen time and fast-food consumption after one year but did not significantly reduce BMI.

These findings are consistent with a broader pattern observed in pediatric obesity prevention research [14]: behavioral counseling interventions, including MI, tend to produce meaningful improvements in parental awareness, feeding practices, and lifestyle routines, yet often yield small or inconsistent effects on BMI trajectories, particularly in very young children [15]. This has led researchers to argue that early prevention may benefit less from intensifying in-person counseling and more from supporting sustained parental reflection, self-regulation, and microbehavioral changes embedded in daily family life [16].

In parallel, digital health interventions have increasingly been explored as a means to provide scalable, low-burden support for families. Systematic reviews of mobile health (mHealth) interventions for childhood obesity prevention have highlighted improvements in engagement, knowledge, and health behaviors, while also emphasizing the importance of co-design, personalization, and integration with existing care pathways [17]. However, many digital tools rely on static content, tracking functions, or gamification, and rarely attempt to replicate the reflective and relational components typical of counseling approaches.

Recent advances in conversational agents and large language models (LLMs) have opened new possibilities for delivering psychologically informed support through dialogue-based interfaces. Conversational agents have shown promise in mental health, chronic disease management, and health behavior change, particularly when designed to simulate elements of empathic communication, guidance, and goal setting [18]. Nevertheless, the use of conversational AI for early childhood obesity prevention remains largely unexplored, and little is known about how core counseling principles can be translated into asynchronous chatbot interactions that take place outside the clinical setting [19].

This gap suggests an opportunity to rethink how family-centered preventive counseling can be delivered: not by replacing professionals, but by extending the space in which reflection and behavior planning can occur, embedding it within everyday routines through accessible, conversational digital tools [20].

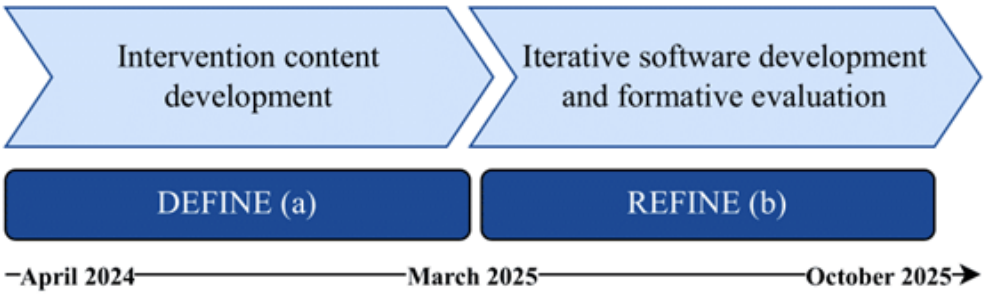
In response to the identified needs, we developed Demetra, a digital health intervention based on the Reggio Emilia program's clinical insights. The core principles of the program are translated into a scalable chatbot-based format. Demetra is delivered via a mobile interface and supports caregivers of children aged 2 to 3 years who are at risk of being overweight (BMI >85th percentile) by providing tailored guidance through conversational dialogue. Demetra is structured around 4 consecutive asynchronous sessions covering values clarification, nutrition, physical activity, and screen-time psychoeducational content and goal setting; the intervention's content, architecture, and use of generative AI are described in detail in The Demetra Intervention section. While the original Reggio Emilia intervention was primarily

focused on MI, Demetra incorporates additional psychological and behavioral change frameworks: Acceptance and Commitment Therapy (ACT) [21] and some key principles of Brief Action Planning (BAP) [22]. These approaches have been selected for their ability to address psychological flexibility, values-based goal setting, and realistic planning. These additions aim to support knowledge, motivation, and the self-regulatory skills that are essential for long-term change in complex family environments.

The Demetra intervention has been developed by a multidisciplinary project team, which included eHealth researchers, experts in psychology and AI. Building on the Reggio Emilia concept introduced previously, the team developed the initial drafts of the intervention and held weekly meetings during the design and development phase. A user-centered design approach was adopted, ensuring the continuous involvement of end users (families), pediatricians, and security specialists, all of whom participated in meetings as needed.

The present study is positioned within phases Ia ("Define") and Ib ("Refine") of the Obesity-Related Behavioral Intervention Trials (ORBIT) model [23], which guides the iterative development of scalable evidence-based behavioral interventions for preventing and managing chronic diseases. The development process comprised two main steps that have been mapped to the ORBIT model (Figure 1) as follows: (1) designing the intervention content and (2) iteratively implementing the prototype software and conducting formative evaluations. ORBIT focuses specifically on the early, pre-efficacy stages through stakeholder-driven design and refinement. In this way, researchers can optimize the intervention, and only when it progresses to clinical trials. Accordingly, the overarching aim of this work was to co-design and conduct a preliminary stakeholder evaluation of Demetra.

Figure 1. Development and iterative refinement of the Demetra intervention (phase I: ORBIT mode design). ORBIT: Obesity-Related Behavioral Intervention Trials.



Specific objectives were to:

1. Iteratively develop and refine the intervention content and a high-fidelity prototype through a multidisciplinary, user-centered co-design process with stakeholders.
2. Evaluate the initial acceptability, engagement, usability, and qualitative perceptions based on prototype interactions involving diverse groups of stakeholders (pediatricians, psychologists, communication/AI experts, and volunteer families).
3. Use stakeholder feedback to improve the content and perform technical optimizations for further proof-of-concept testing in the ORBIT phase IIa.

This work lays the groundwork for future efficacy evaluations and provides practical insights for other conversational AI interventions aimed at preventing early childhood obesity.

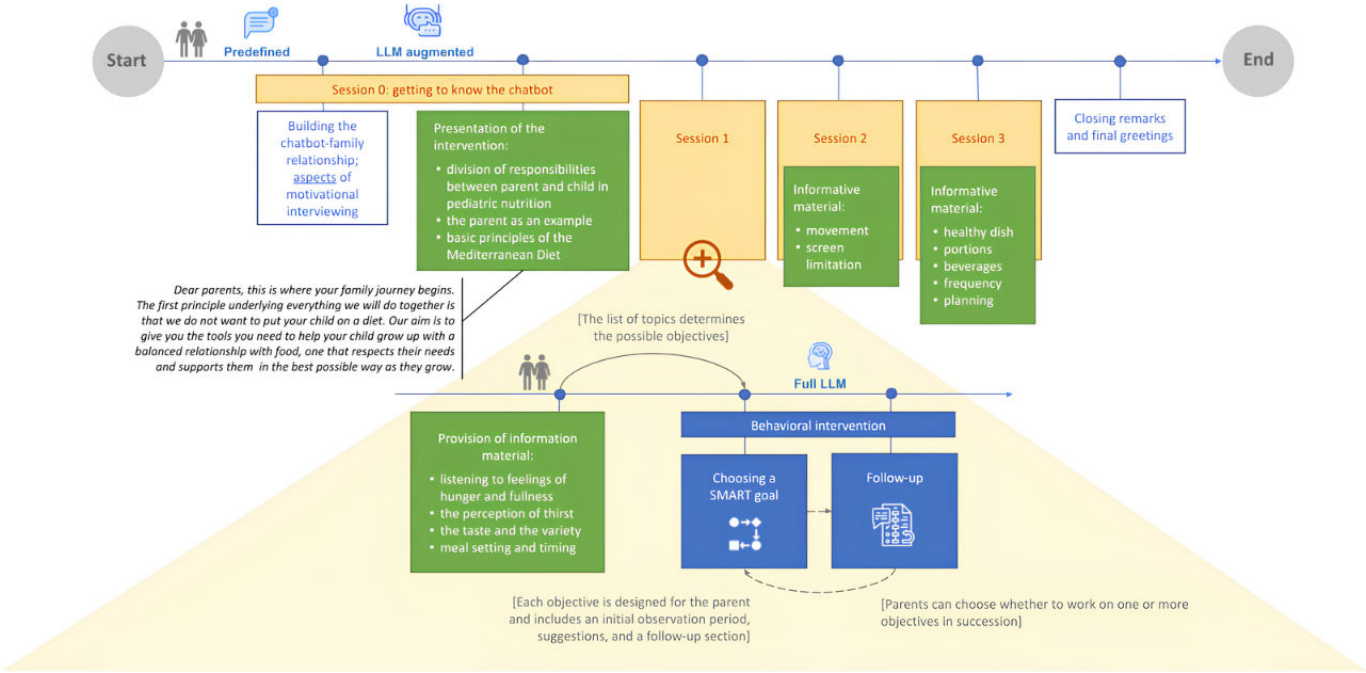
Methods

The Demetra Intervention

Structure and Description

Demetra is an Italian-language mobile app structured around 4 consecutive chatbot sessions (Figure 2). The content has been designed to be accessed asynchronously, thus accommodating families' daily routines without requiring synchronous appointments or technical proficiency and offering flexibility in both timing and pacing.

Figure 2. General structure of the Demetra intervention. LLM: large language model; SMART: specific, measurable, achievable, relevant, and time-bound.



Demetra is aimed at those who are primarily responsible for meals in the family (hereinafter referred to as the caregiver). Each session is meticulously designed to simulate the type of conversation you would have with a professional. It systematically guides users through a reflection and planning process while providing pertinent, informative content. The chatbot uses a combination of scripted and generative AI-driven dialogue (see the Conversational Architecture and Use of Generative AI in Demetra section for more details), informed by behavior change theory. The dialogue is reviewed for linguistic tone, cultural relevance, and psychological appropriateness.

The initial session, designated as "Getting to know the chatbot," comprises an introduction and an orientation module with a focus on establishing a collaborative relationship between the chatbot and the caregiver. The use of MI techniques during the session facilitates the elicitation of the caregiver's values, the examination of family dynamics, and the clarification of motivations for participating in the program. During this phase, the chatbot acquires information regarding the family structure, how much time they spend together, the values that guide parenting choices, and their current perceptions of their children's health. The goal of the session is to build trust and engagement, as well as to promote a process of self-reflection that aligns with the principles of ACT (eg, engaging, focusing, evoking, and planning).

Each of the following 3 sessions addresses a specific thematic domain related to healthy lifestyle behaviors concerning nutrition and movement. Topics include recognizing hunger and satiety cues, physical activity, regulating screen time, and achieving dietary diversity. Each session integrates psychoeducational content, conveyed through brief texts, images, and podcasts, with interactive components designed to promote change. In these components, caregivers are invited to observe their children's specific behaviors in their daily routines and subsequently define concrete behavioral goals (Figure 3A).

Figure 3. Screenshots of the Demetra intervention: (A) findings from the observation period proposed to the caregiver by Demetra, (B) caregiver formulating a SMART (specific, measurable, achievable, relevant, and time-bound) objective with Demetra's support, and (C) follow-up session.



(A) (B) (C)

The goal-setting process is guided using a BAP-inspired module that helps caregivers define specific, realistic, and actionable steps (SMART [specific, measurable, achievable, relevant, and time-bound] objectives) that align with their values and readiness to change (Figure 3B). The chatbot has been developed to facilitate this process by prompting users to identify potential facilitators and barriers. In order to promote continuity and accountability, the chatbot follows up with the user 1 to 2 weeks after each goal-setting interaction to check on progress and encourage the adjustment or reinforcement of goals as required (Figure 3C).

The fourth and final session focuses on reflection and closure. It invites users to review the goals they pursued, reflect on their progress, and develop strategies to turn new behaviors into habits, knowing that it may take more time. This session maintains a supportive and nondirective tone, fostering a sense of autonomy and self-efficacy.

Conversational Architecture and Use of Generative AI in Demetra

Demetra's distinguishing characteristic is the hybrid conversational architecture, which combines scripted content with generative AI to enhance personalization, responsiveness, and emotional resonance. The system achieves a balance between the clinical rigor of theory-driven dialogue and the flexibility and immediacy of LLMs.

Each session is meticulously structured around preauthored dialogue sequences written and reviewed by a multidisciplinary team of clinicians, nutritionists, psychologists, and digital health designers. These scripted segments are designed to deliver the primary psychoeducational content while ensuring consistent alignment with behavior change models, therapeutic tone, and cultural appropriateness. Scripting also enables careful control over language and sequencing, which is essential when guiding users through reflective or emotionally sensitive material.

To enhance user engagement and increase the chatbot's perceived empathy, Demetra uses LLM-augmented responses [24] in selected single-turn interactions. In such cases, OpenAI's GPT-4.1o mini model is used to generate context-sensitive responses based on user input (Figure 4A). These augmentations are tightly constrained through prompt engineering and postprocessing filters to ensure content safety and therapeutic congruence. For content safety, the postprocessing filter uses Llama Guard 3 8B as a downstream classifier, with custom safety categories defined for the intervention; responses classified as potentially harmful are regenerated until they are classified as safe. The purpose of these AI-assisted responses is not to introduce new content, but rather to enhance the emotional depth of the conversation by tailoring the responses to the caregiver's unique language and experiences [24]. The usability study also investigated

and explored in depth the participants' possible perception of bias or incorrect/fictitious responses provided by Demetra.

Figure 4. Screenshots of the Demetra intervention: (A) large language model (LLM)–augmented responses and (B) a multiturn conversation in a follow-up session.



The most dynamic use of generative AI occurs in the BAP modules, which are entirely driven by a LLM. In these segments, the chatbot engages users in multiturn conversations that support collaborative goal setting (Figure 4B). The sections are implemented using a role-playing prompting template [25] wherein an agent is instructed to simulate the logical flow of a BAP session [26]. This includes eliciting a concrete behavioral goal, exploring readiness and motivation, identifying practical steps, and discussing potential facilitators or barriers. Instructions following the capabilities of the agent have been evaluated during phase Ia of ORBIT using a simulation tool called VALISE [27].

Participants

A total of 32 adult volunteers (women and men) were recruited in Italy through professional networks and word-of-mouth sampling to ensure the inclusion of different professional perspectives relevant to the prevention of early childhood obesity and the design of digital health interventions. Participants represented 4 stakeholder groups: pediatricians (working in primary care, private practice, or hospital settings), psychologists, communication and AI experts, and families (caregivers of young children).

Eligibility criteria included being aged at least 18 years and providing informed consent. Families were not required to have a child with a BMI above the 85th percentile or a diagnosis of overweight. Including children identified as being at risk of overweight would have required additional ethical and legal procedures that were beyond the scope of this early-stage research. No additional exclusion criteria were applied, other than an inability to consent or insufficient Italian language proficiency to interact with the chatbot content. None of the participants involved in the evaluation had previously participated in the co-design meetings.

Measures

In accordance with the established practices [28-30] for the development of early-phase behavioral interventions within the ORBIT framework, a user-centered, mixed methods convergent parallel design [31] was adopted for the purpose of obtaining multidisciplinary stakeholder feedback on a high-fidelity prototype of the Demetra intervention deployed on Telegram. A combination of quantitative

questionnaires and qualitative semistructured interviews was used to capture complementary perspectives on acceptability, usability, engagement, educational value, and perceptions of AI-mediated interaction. Quantitative measures, reported in Table 1 with full instruments in Multimedia Appendix 1, provided structured outcome data; the qualitative component explored participants' in-depth experiences with personalization, perceived empathy, conversational safety, cultural and linguistic appropriateness, usability facilitators and barriers, and priorities for refinement.

Table 1. Quantitative instruments used to evaluate acceptability, usability, conversational quality, and educational media in the Demetra prototype.

Instrument	Description	Response scale	Reference
Semantic differential	Item set adapted from a study evaluating a digital porting of a stress-related intervention; assesses perceived empathy and listening skills, conversational flow and fluency, overall chatbot interaction quality, and perceived appropriateness of single-session duration	5-point bipolar adjective pairs	[28]
uMARS[a]	Selected subscales of the uMARS assessing engagement (section A), information quality (section D), subjective quality (section E), and perceived impact (section F); domains were chosen for relevance to high-fidelity prototype evaluation	5-point Likert	[32]
CES-LCC[b]	Assesses chatbot capabilities across understanding user requests, providing helpful information, clarity and relevance of responses, language quality, trust, emotional support, guidance and direction, memory, and overall satisfaction	5-point Likert	[33]
Ad hoc podcast evaluation items	Evaluates podcast quality across 4 conceptual dimensions: content value (informativeness, information quality, topic relevance, and perceived usefulness and impact), engagement (interest and motivation, and net promoter score), presentation quality (monologue format, narrator voice, and background music), and structural design (content frequency and episode duration)	5-point Likert	This study
Ad hoc educational media evaluation items	Evaluates nonaudio educational materials across 4 conceptual dimensions: content value (perceived usefulness, relevance, and fulfillment of expectations), engagement (interest and motivation), presentation quality (clarity, comprehensibility, organization of information, accessibility, and format adequacy), and structural design (usability, navigation, and appropriateness of time required for consultation)	5-point Likert	This study

[a]uMARS: user version of the Mobile Application Rating Scale.
[b]CES-LCC: Comprehensive Evaluation Scale for LLM-Powered Counseling Chatbots.

Ethical Considerations

The study protocol was reviewed and approved by the Research Ethics Committee of the University of Trento (protocol number 2025-057). All procedures complied with institutional ethical standards and with the principles of the Declaration of Helsinki. Participants received detailed study information, a privacy notice, and an electronic informed consent form before participation.

Data collection followed a sequential, standardized procedure:

- Initial questionnaire: Participants completed an online questionnaire that collected demographic information (eg, age, gender, and educational level), stakeholder group, trust in technology (as measured by [34]), self-reported familiarity with chatbots and LLMs, podcast listening habits, and prior experience with mHealth apps.
- Prototype interaction: Participants were granted access to the Demetra Telegram prototype and asked to complete the intervention at least 3 times, adopting a different fictional user profile in each run: an uncooperative user, an ideal user, and an average user with an overweight child. Each run covered the full intervention, which extends over multiple days, allowing the chatbot's conversational robustness, adaptability, and safety to be assessed across different user profiles.
- Quantitative survey: After completing the prototype interactions, participants filled in the quantitative evaluation questionnaires.
- Qualitative interview: Finally, brief semistructured interviews were conducted either in person or via Google Meet. The interviews were audio-recorded to allow for accurate transcription and analysis.

Participants who completed all study procedures received a small nonmonetary token of appreciation, such as a reusable water bottle, pen, or similar item. All interviews were transcribed verbatim and, together with quantitative study data, pseudonymized. All data were stored securely and handled in accordance with applicable data protection regulations.

Data Analysis

In view of the exploratory nature of the study and the small, convenient sample size, the quantitative data were primarily summarized using descriptive statistics, such as means, SDs, medians, and ranges. To provide a minimal inferential assessment of overall perceptions, one-sample, 1-tailed Wilcoxon signed-rank tests [35] were conducted to evaluate whether the median scores were greater than the neutral scale midpoint (value=3) using an α level of .05. No subgroup, comparative, or multivariable analyses were performed because the sample size was insufficient to support them; therefore, subgroup-related observations are reported descriptively only.

The qualitative interview transcripts were analyzed using a thematic analysis based on the six-phase approach of Braun and Clarke [36]. Two researchers (AEN and BP) reviewed the transcripts independently, discussed the emerging themes,

and resolved discrepancies through consensus. The quantitative and qualitative findings were integrated during the interpretation process to generate convergent insights and actionable recommendations for optimizing the prototype and planning ORBIT phase IIa. For reporting purposes, the qualitative feedback from interviews conducted in Italian was translated into English, and participants were identified using group-specific codes, with numbers assigned sequentially within each stakeholder group (eg, FAMILY2 and PEDIATRICIAN3).

Results

Participant Characteristics and Study Overview

Of the 32 participants who were initially recruited and provided consent, 26 completed all required interactions and assessments, resulting in a dropout rate of 19% (6/32). The dropouts were all professionals, but they were not evenly distributed across the 4 stakeholder groups (3 pediatricians, 2 psychologists, and 1 communication/AI expert). No families withdrew. During follow-up contact, they reported that the only reason for noncompletion was because they could not find enough time to interact with the prototype according to the various proposed scenarios within the study window. This was particularly evident among the pediatricians and psychologists, whose clinical caseloads limited their availability. No participants withdrew due to dissatisfaction with the prototype or technical difficulties.

The final sample (n=26) included participants of various ages, genders, educational levels, stakeholder groups, levels of trust in technology, and levels of self-reported knowledge of LLMs and chatbots, as summarized in Table 2.

Table 2. Participant characteristics by stakeholder group.

Characteristic	Pediatricians	Families	Psychologists	Communication/AI experts
Age (y), mean (SD)	45.43 (14.29)	34.62 (3.54)	31.20 (2.95)	37.17 (11.86)
Gender	0 men, 7 women, and 0 nonbinary people	2 men, 6 women, and 0 nonbinary people	2 men, 3 women, and 0 nonbinary people	4 men, 6 women, and 0 nonbinary people
Educational level, n				
High school or lower	0	2	0	0
Bachelor's degree	0	1	0	1
Master's or higher	7	5	5	5
Propensity to trust technology (values 1-5), mean (SD)	3.45 (0.25)	3.33 (0.52)	3.40 (0.57)	3.64 (0.25)
LLM[a] experience (values 1-4), mean (SD)	2.00 (0.82)	2.00 (0.93)	3.00 (0.71)	3.50 (0.55)
Chatbot experience (values 1-4), mean (SD)	1.86 (0.69)	1.75 (0.71)	2.20 (0.45)	3.17 (0.75)

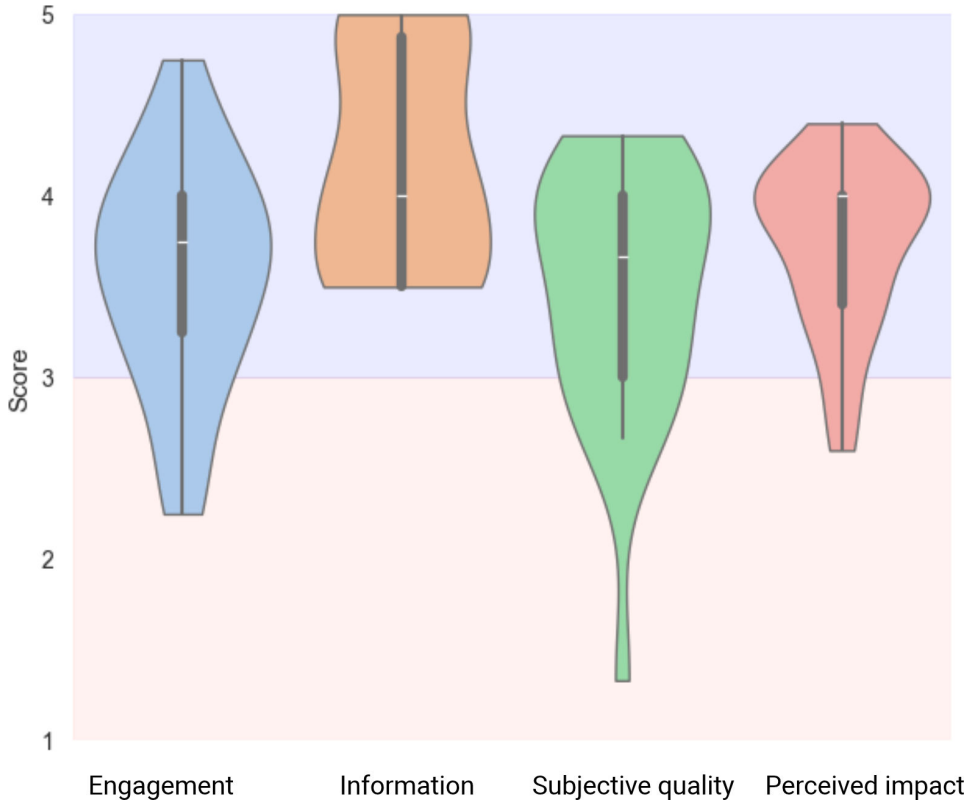
[a]LLM: large language model.

Quantitative Results

Overall Prototype Quality and Perceived Impact

The prototype was evaluated positively across all selected uMARS domains, with median scores consistently and significantly exceeding the neutral midpoint of the scale (3). As shown in Figure 5, the highest ratings were observed for information quality (median 4.00, IQR 3.50-4.88; V=351; P<.001) and perceived impact (median 4.00, IQR 3.40-4.00; V=318; P<.001), indicating strong evaluations of the content and its potential usefulness.

Figure 5. The user version of the Mobile Application Rating Scale (uMARS) domain scores for the Demetra prototype. The blue shaded areas indicate ratings that are above the acceptable level, and the red shaded areas indicate ratings that are below the acceptable level.



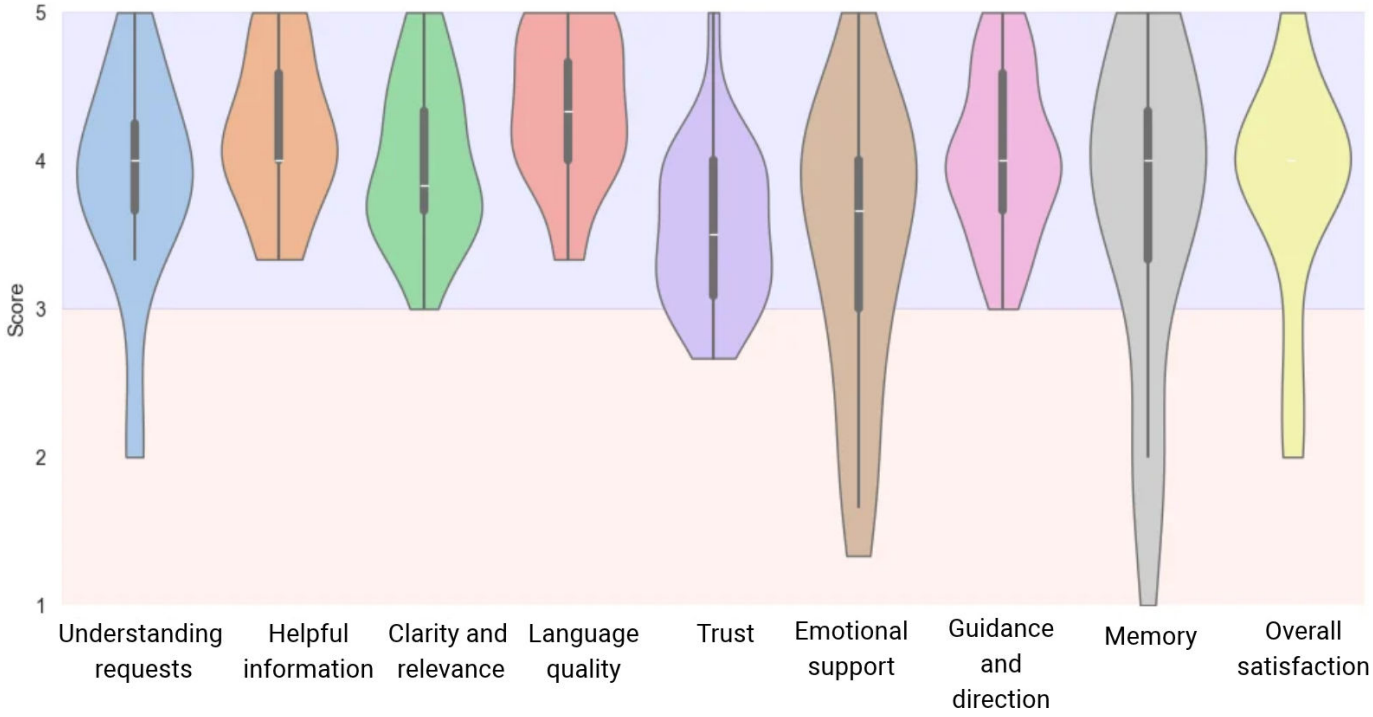
Although engagement (median 3.75, IQR 3.25-4.00; V=252; P<.001) and subjective quality (median 3.67, IQR 3.00-4.00; V=219; P<.001) received slightly lower median scores, both domains remained significantly above the midpoint. Responses in these areas, however, displayed greater variability, suggesting differences in individual experiences and expectations. Nonetheless, the overall pattern across domains points to broad acceptability of the prototype at this stage of development.

Conversational Quality, Trust, and Empathy

Participants gave Demetra positive evaluations across both the CES-LCC and semantic differential measures. All relevant dimensions were rated significantly above the neutral scale midpoint.

As illustrated in Figure 6, the strongest aspects of the interaction were the language quality, clarity, and relevance of the responses, and the guidance and direction provided, indicating that users perceived the chatbot as understandable, well-structured, and supportive in guiding them through reflection and goal-setting processes. Ratings for memory and emotional support were moderately positive, though they showed greater variability among individuals, reflecting potential differences in expectations for AI-mediated relational support.

Figure 6. Comprehensive Evaluation Scale for LLM-Powered Counseling Chatbots (CES-LCC) domain scores for the Demetra prototype. The blue shaded areas indicate ratings that are above the acceptable level, and the red shaded areas indicate ratings that are below the acceptable level.



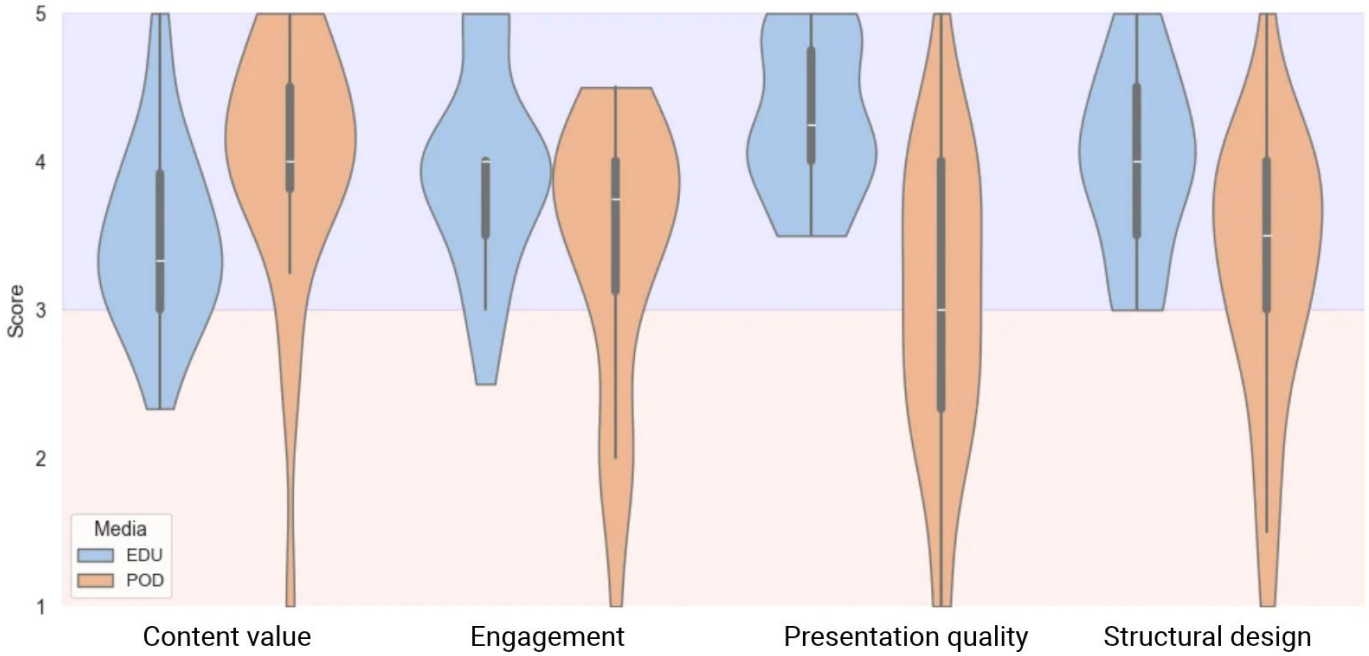
The results of the semantic differential further supported these findings. The interaction was evaluated by the participants as empathic and fluent with ratings for empathy and listening (median 3.80, IQR 3.65-4.35; *V*=294; *P*<.001) and conversational flow and fluency significantly above the midpoint (median 4.00, IQR 3.25-4.00; *V*=192; *P*<.001). The median rating for the overall quality of the interaction with the chatbot was also above the midpoint (median 3.50, IQR 3.00-3.67; *V*=219; *P*=.02).

The single interaction duration dimension (median 3.00, IQR 2.50-3.50; *V*=103; *P*=.38) was the only item of the semantic differential for which scores did not differ significantly from the scale midpoint. Importantly, the midpoint value (3) for this item represents the ideal duration rather than a neutral or undesirable condition. Therefore, the absence of deviation from the midpoint indicates that the length of individual interactions was perceived as appropriate and well-balanced, neither too short nor overly demanding.

Evaluation of Podcasts and Nonaudio Educational Media

As shown in Figure 7, the content value of podcasts was rated significantly above the midpoint (median 4.00, IQR 3.81-4.50; *V*=322; *P*<.001), indicating that participants found the information to be relevant, understandable, and potentially useful. Engagement was also rated above the midpoint (median 3.75, IQR 3.12-4.00; *V*=220; *P*=.02), though there was greater variability among participants. In contrast, presentation quality (median 3.00, IQR 2.33-4.00; *V*=144; *P*=.28) and structural design (median 3.50, IQR 3.00-4.00; *V*=145; *P*=.07) did not differ significantly from the midpoint. This suggests that aspects such as format, pacing, and episode structure were perceived as acceptable but not optimal.

Figure 7. Participant evaluation of podcasts and nonaudio educational materials. The blue shaded areas indicate ratings that are above the acceptable level, and the red shaded areas indicate ratings that are below the acceptable level. EDU: nonaudio educational media; POD: podcast.



In contrast, nonaudio educational materials showed a more consistently positive profile. The median ratings for content value (median 3.62, IQR 3.00-3.92; V=348; P<.001), engagement (median 4.00, IQR 3.50-4.00; V=273; P<.001), presentation quality (median 4.00, IQR 4.00-4.75; V=325; P<.001), and structural design (median 4.00, IQR 3.50-4.50; V=253; P<.001) were all significantly above the midpoint. These results indicate that the educational materials were perceived as clear, accessible, and well-suited to support asynchronous interaction with the chatbot.

Extended quantitative results are available in Multimedia Appendix 2.

Qualitative Results

The qualitative analysis of the semistructured interviews revealed 4 overarching themes that transcend stakeholder groups. The analyses also revealed differences in emphasis related to professional role and lived experience. The themes concern (1) the clarity, usability, and tone of the chatbot's communication; (2) the perceived relevance and adequacy of the content for families with young children; (3) the engagement, motivation, and effort required for sustained use; and (4) the perceived role of professionals and the limits of AI-mediated support.

Clarity, Usability, and Communicative Tone

Participants from all stakeholder groups repeatedly emphasized the clarity of Demetra's language and how easy it was to use the application. In particular, caregivers and pediatricians valued the accessibility of the content and the flexibility of engagement, which did not require technical skills or prolonged attention.

Several caregivers emphasized that the linguistic simplicity and mobile delivery supported its use in everyday situations marked by time constraints and multitasking:

> *I really appreciated the simplicity of the language; it's very easy to understand, and obviously the fact that you can listen to things and listen again whenever you want. You can interrupt, pick it up again later… and of course, have access on your phone. The podcasts are something you can even listen to while driving.* [FAMILY2]

Pediatricians echoed this perception of ease and flexibility, framing clarity and brevity as prerequisites for the effectiveness of preventive interventions:

> *It felt like an attentive and quick interaction, which is exactly what I like. Interventions that are effective and straight to the point.* [PEDIATRICIAN4]

Psychologists and communication experts focused more explicitly on the quality of conversation, noting that it was generally perceived as fluid and coherent. In some cases, it was even considered to be surprisingly close to a human exchange. Several participants also remarked on the chatbot's ability to refer back to user inputs and maintain coherence within the interaction, which contributed to the feeling of being listened to:

> *The language was appropriate, it picked up on what I had said before, it referred back to my answers, and that made you feel listened to.* [PSYCHOLOGIST3]

At the same time, this generally positive evaluation was accompanied by more nuanced observations. Some participants, particularly psychologists and communication

experts, noted that, although the tone was consistently supportive and nonjudgmental, it could sometimes seem overly accommodating or formulaic, particularly during repeated interactions. These remarks did not negate perceptions of clarity or usability, but introduced early signals of tension between accessibility, personalization, and conversational depth, which reemerge in later themes.

Perceived Relevance and Adequacy of Content for Families With Young Children

Participants largely agreed that Demetra addresses a developmentally significant phase of early childhood, and that focusing on caregivers of children aged 2 to 3 years is appropriate. Pediatricians emphasized the importance of intervening during this period, when eating habits, routines, and parental practices begin to consolidate:

> *In my opinion this is the most important age range, because this is where food education really begins. What you learn at this age is what you tend to carry with you later on.* [PEDIATRICIAN6]

From a clinical perspective, pediatricians also valued the scientific coherence of the messages and their alignment with established nutritional and behavioral recommendations of the messages conveyed by the chatbot, noting that it did not convey misleading or inaccurate information:

> *The information was certainly good and accurate. It's not like it was making things up or saying anything that wasn't correct.* [PEDIATRICIAN3]

Similarly, caregivers recognized the relevance of the topics addressed and often described small but concrete attempts to modify daily routines after interacting with the chatbot. These accounts suggest that the content was perceived as connected to everyday practices rather than abstract:

> *Three weeks isn't much, of course, but I did try to reduce certain things, like snacks. I tried to work on the part that I felt was more critical.* [FAMILY2]

On the other hand, caregivers also expressed a desire for situation-specific guidance, especially in complex family situations. For parents managing multiple children with different needs, Demetra's recommendations were sometimes difficult to implement:

> *What's missing, maybe, is something more practical, more targeted advice for the parent herself about food. Especially when you have two children with different needs and different timings, it becomes really hard to stick [even] to good intentions.* [FAMILY2]

Some pediatricians echoed the request for more concrete information, suggesting that providing examples, recipes, or simple demonstrations could make the information more applicable.

Engagement, Motivation, and Effort Required for Sustained Use

Participants described their engagement with Demetra as closely tied to their available time, personal motivation, and ability to integrate the interaction into their daily routines. Rather than viewing engagement as an inherent property of the intervention alone, many participants emphasized the effort required on the user's part and the conditions under which continued use felt feasible.

Caregivers often portrayed Demetra as a potentially supportive presence. However, they also acknowledged that sustained engagement depended on their own readiness and capacity to invest attention over time. One participant stated:

> *It uses empathy to acknowledge you and then moves on to the next step.* [FAMILY4]

Differences in usage patterns also emerged. Some participants chose to space interactions over several days, while others preferred to complete longer sessions in a single sitting. These differences suggest heterogeneous strategies for managing cognitive and temporal load. Despite this flexibility, several participants described difficulties related to continuity and follow-up. In particular, they found that the responsibility of remembering to return to the intervention and monitor progress was perceived as demanding, especially in already busy family contexts:

> *It's hard to dedicate time and attention. You lose moments.* [FAMILY7]

> *It's difficult to remember and to take on that responsibility yourself.* [PSYCHOLOGIST3]

From a professional perspective, pediatricians and psychologists interpreted these challenges as intrinsic to preventive interventions that rely on self-regulation, rather than as straightforward usability issues. However, communication experts also noted that repetitive interaction patterns and limited variation in the conversational flow could undermine engagement over time:

> *The repetitiveness of the approach can be a barrier… in the long term I'm afraid [the intervention] might not be very engaging… it could lack elements of novelty.* [UAIXEXPERT2]

Role of Professionals and Perceived Limits of AI-Mediated Support

Demetra was consistently framed as a supportive tool rather than a replacement for professional care across stakeholder groups. Pediatricians, psychologists, and caregivers converged on the idea that the chatbot could complement existing clinical or educational pathways, particularly by providing continuity and reinforcement between in-person encounters.

Pediatricians emphasized the potential value of Demetra for families with fewer resources or less access to structured support. However, they maintained that professional oversight remains essential:

> *It can work and it can help people who are really a bit on their own… but they still need to be followed by the pediatrician*. [PEDIATRICIAN7]

Similarly, caregivers described the chatbot as a supportive presence that could provide reassurance and continuity outside of clinical settings, especially between scheduled appointments with health care providers. They framed this sense of support "in between" especially when professional guidance was unavailable, while clearly distinguishing it from direct clinical care.

Several professionals, particularly psychologists and communication experts, raised concerns about the limitations of AI-mediated interactions in addressing the emotional complexities of parenting. One recurring issue was the chatbot's tendency to maintain a uniformly positive and encouraging tone, even when users expressed difficulty, fatigue, or vulnerability. This approach was seen as not responsive enough to nuanced emotional states and as potentially misaligned with distressing situations:

> *It didn't really understand what I was saying, it was too focused on strengths. I'm fine, but what if I weren't? It completely misses people's difficulties*. [PSYCHOLOGIST1]

> *My concern is that it seems to go along too much with what the user says. It doesn't really analyze or contextualize the answers*. [AICOMEXPERT3]

These reflections delineate the boundary between what participants considered appropriate for an AI-based intervention and what they believed required human judgment, emotional attunement, and professional responsibility. Although Demetra was considered capable of providing structure, information, and motivational support, its role was clearly limited within a broader care ecosystem in which professionals play a central role.

An overview of qualitative results is available in Multimedia Appendix 3.

Discussion

Principal Findings

We developed Demetra in response to a practical limitation of early-childhood overweight prevention: while reflective, family-centered approaches are considered appropriate and potentially effective at young ages, they are difficult to sustain through repeated in-person encounters. This difficulty has been widely described in the literature on pediatric MI, where behavioral improvements are often observed but scalability and long-term implementation remain challenging due to time constraints in primary care and the need for repeated professional involvement [37-39]. The feedback gathered in this study indicates that families and professionals consider it reasonable to introduce this type of support during the 2- to 3-year age range. This aligns with evidence highlighting early childhood as a sensitive period for the establishment of eating routines, self-regulation, and parental feeding practices, where preventive efforts may have long-lasting influence even when BMI changes are not immediately detectable [40-42].

In this study, participants often described the interaction as compatible with everyday life, suggesting that preventive work at this stage may benefit from formats that allow parents to engage at their own pace rather than within the constraints of scheduled visits. This finding resonates with previous research on mHealth interventions for childhood obesity prevention, which emphasizes flexibility, low burden, and integration into daily routines as key factors for engagement and feasibility [17,43,44].

A central question for us was whether the core elements of the Reggio Emilia MI program [10] could be preserved when removed from the pediatric consultation. Stakeholders consistently recognized familiar elements, including attention to parental values, an emphasis on small and realistic goals, and nonadjunct tools that extend care rather than use a directive tone. This is particularly relevant in light of prior studies showing that MI-based pediatric interventions often succeed in improving parental awareness, feeding practices, and lifestyle routines, even when BMI outcomes remain modest [13-15]. The main difference concerned where and when this work occurred. In Demetra, reflection and planning occurred in domestic settings, and were often spread across short periods of time, rather than concentrated in a single clinical encounter. Pediatricians in particular interpreted this shift as a way to extend preventive care beyond the visit without increasing their workload, while maintaining coherence with their usual counseling approach. This interpretation is consistent with calls in the literature to move from episodic counseling toward sustained behavioral support embedded in family contexts [45,46]. Importantly, the positive ratings on conversational quality, empathy, and guidance suggest that key relational components of counseling, typically considered difficult to reproduce digitally, were at least partially perceived as present. This finding is coherent with emerging evidence on conversational agents in health care, where users report feelings of being "heard" and supported when dialogue systems are carefully designed to simulate empathic communication and goal-oriented guidance [18,19,47].

Indeed, Demetra was never intended to replace professional care. Parents, pediatricians, and psychologists emphasized the importance of supervision and the limitations of digital tools, particularly when working with families of very young children. These observations mirror concerns raised in the literature on AI-mediated health support, where chatbots are increasingly framed as adjunct tools that extend care rather than substituting for human judgment, especially in emotionally complex contexts [48,49]. Based on this feedback, we

see Demetra as fitting most naturally within a stepped care approach to prevention [50]. Pediatricians could introduce the tool during routine visits and later use follow-up appointments to discuss what parents tried, what difficulties emerged, and whether additional support is warranted. This hybrid integration between digital support and professional follow-up has been repeatedly identified as a promising model for digital behavior change interventions [51,52].

Limitations and Future Research

At this stage, we examine how families and professionals understood, used, and positioned the approach with high-fidelity prototypes, rather than testing a finished version of the intervention. This approach is consistent with the ORBIT model's emphasis on iterative refinement before efficacy testing, and with recommendations in digital health research that stress the importance of usability and acceptability assessments before outcome evaluation [23,53].

The study relied on a relatively small, convenience sample and was not designed to capture the full range of family contexts encountered in routine pediatric practice. Participants were recruited through professional networks and word-of-mouth, which may have limited the socioeconomic diversity of the sample, particularly by underrepresenting families less connected to these networks. In addition, family participants were not selected based on their child's age or BMI, and therefore do not necessarily reflect the specific population ultimately targeted by Demetra. This is relevant given evidence that digital health engagement varies substantially across socioeconomic and educational groups, and that families with fewer resources may face greater barriers to sustained use [54,55]. The sample also included some imbalances across subgroups (eg, no male pediatricians), and participants' professional proximity to the context in which Demetra was developed may have influenced their responses, including through socially desirable reporting. Participants also interacted with Demetra for a limited time and under structured conditions. The interactions were based on fictitious scenarios rather than on participants' own experiences. The average-user scenario was designed to approximate the type of interaction expected in routine use, while the other scenarios were included to examine how Demetra responded to interactions that departed from this expected condition. This approach allowed us to explore a broader range of possible interactions, but limits the interpretation of responses concerning empathy, personal relevance, and perceived usefulness, particularly where these depend on users' own circumstances. Moreover, while this made it possible to explore initial reactions, perceived fit, and plausible modes of use, it does not inform how the tool would be incorporated into everyday routines over time or how engagement might fluctuate. This issue reflects a broader challenge documented in the mHealth literature, where early positive engagement often declines over time unless specific strategies for novelty, reminders, and personalization are implemented [56,57]. Several qualitative findings, such as requests for more practical examples, greater contextual specificity, and more nuanced emotional responses, also highlight known limitations of conversational agents in health behavior change. Prior studies have shown that users may initially appreciate supportive tone and clarity but later perceive interactions as repetitive or overly generic if adaptive variation is limited [58].

Based on these findings, the next step is a proof-of-concept study that corresponds to the ORBIT phase IIa. During this phase, Demetra will be implemented as a dedicated application that incorporates the refinements identified through the evaluation. The phase IIa study will examine the feasibility of integrating the revised version into pediatric workflows, and engagement patterns focusing on proximal processes relevant to early prevention, such as parental awareness, motivation, and follow-through, before assessing downstream health outcomes. This progression is coherent with recommendations to evaluate behavioral mediators before BMI outcomes in early childhood prevention research [59,60].

Future work should also assess the computational and financial costs associated with sustained use of generative AI at scale, including whether smaller and more cost-efficient models can provide sufficiently effective performance for the intended functions.

Finally, Demetra and all of its informative content were developed in the Italian language for the Italian context. Further research is necessary to make the chatbot accessible to other communities living in the province, such as those from Romania, Albania, and Morocco who have historically been most deeply rooted in the area. It is not simply a matter of translation; it is also a matter of culturally adapting the intervention where necessary. Cultural mediators could help to identify areas for intervention, such as linguistic, cultural, pragmatic, ethical, and safety-related issues, and to develop a collaboration model.

Conclusions

The present study delineates the design and early stakeholder evaluation of Demetra, a high-fidelity prototype for the prevention of early childhood overweight, through an asynchronous, chatbot-mediated intervention for caregivers of children aged 2 to 3 years. Inspired by clinical insights from the program in Reggio Emilia, Demetra explores whether core counseling principles, such as reflection, values clarification, and realistic goal setting, can be translated into a scalable digital format that fits naturally within everyday family routines during a sensitive developmental period.

Stakeholder feedback suggests that these principles remain recognizable and acceptable within the chatbot interaction. Participants particularly valued how the intervention supports reflection and practical goal setting in the domestic context while maintaining coherence with pediatric care. At the same time, the evaluation highlighted important limitations of AI-mediated support, especially regarding emotional nuance, contextual specificity, and adaptive conversational depth. These findings reinforce the need to position Demetra as a complement to, rather than a replacement for, professional care.

By integrating co-design and formative evaluation within ORBIT phase I, this study allowed key design tensions and implementation constraints to emerge before efficacy testing. While the study does not provide evidence of effectiveness, it demonstrates how an intervention inspired by an established counseling model can be reshaped for scalability without losing clinical coherence. These results provide a grounded rationale for progressing to ORBIT phase IIa, where a refined version of Demetra will be evaluated for feasibility, sustained engagement, and integration within routine pediatric prevention pathways.

Acknowledgments

The authors would also like to thank all the experts who participated in the co-design process. Their contributions were essential to improving the prototype for the next iteration of the Obesity-Related Behavioral Intervention Trials model.

Generative AI Use

The authors used ChatGPT (OpenAI) and DeepL Write to improve the readability and language of this article during its preparation. After using these tools, they reviewed and edited the content as needed, taking full responsibility for the publication's content.

Funding

This work was funded by Life Science Hub – Digital Health (PNC-E3-2022-23683267 - DHEAL-COM Project– CUP: C63C22001970001), Ministry of Health, as part of the National Complementary Plan for Innovative Health Ecosystems. This publication reflects only the authors' view, and the Italian Ministry of Health is not responsible for any use that may be made of the information it contains.

Data Availability

An anonymized, privacy-friendly version of the dataset, identified only by participant codes, together with the scripts used for quantitative analyses, is included as part of the study materials in Multimedia Appendix 4.

Authors' Contributions

BP, AEN, MB, SDC, LD, SR, SP, CDL, SS, and RF all made substantial contributions to conceptualization, methodology, intervention design, and writing and participated in the co-design and refinement of the Demetra intervention. All authors contributed to data curation, formal analysis, interpretation of quantitative and qualitative data, and writing review and editing and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Questionnaire details and qualitative interview trace.
[DOCX File (Microsoft Word File), 569 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Extended quantitative results.
[ZIP File (ZIP archive File), 231 KB-Multimedia Appendix 2]

Multimedia Appendix 3

Overview of qualitative results.
[DOCX File (Microsoft Word File), 10 KB-Multimedia Appendix 3]

Multimedia Appendix 4

Data and script for analysis.
[DOCX File (Microsoft Word File), 7 KB-Multimedia Appendix 4]

References

1. Reilly JJ, Kelly J. Long-term impact of overweight and obesity in childhood and adolescence on morbidity and premature mortality in adulthood: systematic review. Int J Obes (Lond). Jul 2011;35(7):891-898. [doi: 10.1038/ijo.2010.222] [Medline: 20975725]
2. Rankin J, Matthews L, Cobley S, et al. Psychological consequences of childhood obesity: psychiatric comorbidity and prevention. Adolesc Health Med Ther. 2016;7:125-146. [doi: 10.2147/AHMT.S101631] [Medline: 27881930]
3. Nardone P, Spinelli A, Ciardullo S, Salvatore MA, Andreozzi S, Galeone D. Obesità e stili di vita dei bambini: OKkio alla SALUTE 2019 [Report in Italian]. Istituto Superiore di Sanità; 2022. URL: https://www.iss.it/documents/20126/6682486/22-27%2Bweb.pdf/659d3d23-9236-e3ff-fbbf-0e06360e3e71? [Accessed 2026-11-10]

4. Kartiosuo N, Ramakrishnan R, Lemeshow S, et al. Predicting overweight and obesity in young adulthood from childhood body-mass index: comparison of cutoffs derived from longitudinal and cross-sectional data. Lancet Child Adolesc Health. Nov 2019;3(11):795-802. [doi: 10.1016/S2352-4642(19)30204-4] [Medline: 31451394]
5. Smith JD, St George SM, Prado G. Family-centered positive behavior support interventions in early childhood to prevent obesity. Child Dev. Mar 2017;88(2):427-435. [doi: 10.1111/cdev.12738] [Medline: 28195411]
6. Hesketh KD, Campbell KJ. Interventions to prevent obesity in 0-5 year olds: an updated systematic review of the literature. Obesity (Silver Spring). Feb 2010;18 Suppl 1:S27-S35. [doi: 10.1038/oby.2009.429] [Medline: 20107458]
7. Cueto V, Sanders LM. Telehealth opportunities and challenges for managing pediatric obesity. Pediatr Clin North Am. Aug 2020;67(4):647-654. [doi: 10.1016/j.pcl.2020.04.007] [Medline: 32650862]
8. Kitzmann KM, Beech BM. Family-based interventions for pediatric obesity: methodological and conceptual challenges from family psychology. J Fam Psychol. Jun 2006;20(2):175-189. [doi: 10.1037/0893-3200.20.2.175] [Medline: 16756393]
9. Rollnick S, Miller WR. What is motivational interviewing? Behav Cogn Psychother. Oct 1995;23(4):325-334. [doi: 10.1017/S135246580001643X]
10. Davoli AM, Broccoli S, Bonvicini L, et al. Pediatrician-led motivational interviewing to treat overweight children: an RCT. Pediatrics. Nov 1, 2013;132(5):e1236-e1246. [doi: 10.1542/peds.2013-1738] [Medline: 24144717]
11. Resnicow K, Harris D, Wasserman R, et al. Advances in motivational interviewing for pediatric obesity: results of the brief motivational interviewing to reduce body mass index trial and future directions. Pediatr Clin North Am. Jun 2016;63(3):539-562. [doi: 10.1016/j.pcl.2016.02.008] [Medline: 27261549]
12. Resnicow K, Delacroix E, Sonneville KR, et al. Outcome of BMI2+: motivational interviewing to reduce BMI through primary care AAP PROS practices. Pediatrics. Jan 1, 2024;153(2):e2023062462. [doi: 10.1542/peds.2023-062462] [Medline: 38282541]
13. Taveras EM, Gortmaker SL, Hohman KH, et al. Randomized controlled trial to improve primary care to prevent and manage childhood obesity: the High Five for Kids study. Arch Pediatr Adolesc Med. Aug 2011;165(8):714-722. [doi: 10.1001/archpediatrics.2011.44] [Medline: 21464376]
14. Waters E, de Silva-Sanigorski A, Hall BJ, et al. Interventions for preventing obesity in children. Cochrane Database Syst Rev. Dec 7, 2011(12):CD001871. [doi: 10.1002/14651858.CD001871.pub3] [Medline: 22161367]
15. Oude Luttikhuis H, Baur L, Jansen H, et al. Interventions for treating obesity in children. Cochrane Database Syst Rev. Jan 21, 2009(1):CD001872. [doi: 10.1002/14651858.CD001872.pub2] [Medline: 19160202]
16. Campbell KJ, Hesketh KD. Strategies which aim to positively impact on weight, physical activity, diet and sedentary behaviours in children from zero to five years. A systematic review of the literature. Obes Rev. Jul 2007;8(4):327-338. [doi: 10.1111/j.1467-789X.2006.00305.x] [Medline: 17578382]
17. Schoeppe S, Alley S, Van Lippevelde W, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act. Dec 7, 2016;13(1):127. [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
18. Laranjo L, Dunn AG, Tong HL, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc. Sep 1, 2018;25(9):1248-1258. [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
19. Fulmer R, Joerin A, Gentile B, Lakerink L, Rauws M. Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: randomized controlled trial. JMIR Ment Health. Dec 13, 2018;5(4):e64. [doi: 10.2196/mental.9782] [Medline: 30545815]
20. Hekler EB, Michie S, Pavel M, et al. Advancing models and theories for digital behavior change interventions. Am J Prev Med. Nov 2016;51(5):825-832. [doi: 10.1016/j.amepre.2016.06.013] [Medline: 27745682]
21. Hayes SC, Luoma JB, Bond FW, Masuda A, Lillis J. Acceptance and commitment therapy: model, processes and outcomes. Behav Res Ther. Jan 2006;44(1):1-25. [doi: 10.1016/j.brat.2005.06.006] [Medline: 16300724]
22. Gutnick D, Reims K, Davies C, Gainforth H, Jay M, Cole S. Brief action planning to facilitate behavior change and support patient self-management. J Clin Outcomes Manag. 2014;21(1):17-29. URL: https://www.mdedge.com/jcomjournal/article/147101/endocrinology/brief-action-planning-facilitate-behavior-change-and [Accessed 2026-11-10]
23. Czajkowski SM, Powell LH, Adler N, et al. From ideas to efficacy: the ORBIT model for developing behavioral treatments for chronic diseases. Health Psychol. Oct 2015;34(10):971-982. [doi: 10.1037/hea0000161] [Medline: 25642841]
24. Sanna L, Bolpagni M, Fietta V, et al. LLM-enriched finite-state chatbots for mental health support: a case study on self-help+. In: Bellazzi R, Juarez Herrero JM, Sacchi L, Zupan B, editors. Artificial Intelligence in Medicine: 23rd International Conference, AIME 2025, Pavia, Italy, June 23–26, 2025, Proceedings, Part I. Springer; 2025:386-396. [doi: 10.1007/978-3-031-95838-0_38]

25. Kong A, Zhao S, Chen H, et al. Better zero-shot reasoning with role-play prompting. Proc North Am Chapter Assoc Comput Linguist. 2024:4099-4113. [doi: 10.18653/v1/2024.naacl-long.228]
26. Bolpagni M, De Carli S, Sanna L, Gabrielli S, Dragoni M. Role-play large language models for short behavior change interventions: an exploratory study on brief action planning. In: Bellazzi R, Juarez Herrero JM, Sacchi L, Zupan B, editors. Artificial Intelligence in Medicine 23rd International Conference, AIME 2025, Pavia, Italy, June 23–26, 2025, Proceedings, Part II. Springer; 2025:46-51. [doi: 10.1007/978-3-031-95841-0_9]
27. Bolpagni M, De Carli S, Sanna L, Dragoni M, Gabrielli S. VALISE: a virtual agent laboratory for instruction-following simulation and evaluation of LLM-powered digital health interventions. In: LynceI, MuranoN, VallatiM, VillataS, ChesaniF, MilanoM, OmiciniA, DastaniM, editors. ECAI 2025 – 28th European Conference on Artificial Intelligence, Including 14th Conference on Prestigious Applications of Intelligent Systems (PAIS 2025). IOS Press; 2025:5096-5103. [doi: 10.3233/FAIA251428]
28. Fietta V, Rizzi S, De Luca C, et al. A chatbot-based version of the World Health Organization-validated Self-Help Plus intervention for stress management: co-design and usability testing. JMIR Hum Factors. Oct 18, 2024;11:e64614. [doi: 10.2196/64614] [Medline: 39355954]
29. Graham AK, Munson SA, Reddy M, et al. Integrating user-centered design and behavioral science to design a mobile intervention for obesity and binge eating: mixed methods analysis. JMIR Form Res. May 10, 2021;5(5):e23809. [doi: 10.2196/23809] [Medline: 33970114]
30. Korpershoek YJG, Hermsen S, Schoonhoven L, Schuurmans MJ, Trappenburg JCA. User-centered design of a mobile health intervention to enhance exacerbation-related self-management in patients with chronic obstructive pulmonary disease (Copilot): mixed methods study. J Med Internet Res. Jun 15, 2020;22(6):e15449. [doi: 10.2196/15449] [Medline: 32538793]
31. Creswell JW. Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. SAGE Publications; 2003. ISBN: 9780761924418
32. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the Mobile Application Rating Scale (uMARS). JMIR mHealth uHealth. Jun 10, 2016;4(2):e72. [doi: 10.2196/mhealth.5849] [Medline: 27287964]
33. Bolpagni M, Gabrielli S. Development of a comprehensive evaluation scale for LLM-powered counseling chatbots (CES-LCC) using the eDelphi method. Informatics. 2025;12(1):33. [doi: 10.3390/informatics12010033]
34. Jessup SA, Schneider TR, Alarcon GM, Ryan TJ, Capiola A. The measurement of the propensity to trust automation. In: Chen JYC, Fragomeni G, editors. Virtual, Augmented and Mixed Reality. Applications and Case Studies 11th International Conference, VAMR 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26–31, 2019, Proceedings, Part II. Springer; 2019:476-489. [doi: 10.1007/978-3-030-21565-1_32]
35. Wilcoxon F. Individual comparisons by ranking methods. Biom Bull. Dec 1945;1(6):80-83. [doi: 10.2307/3001968]
36. Clarke V, Braun V. Thematic analysis. J Posit Psychol. May 4, 2017;12(3):297-298. [doi: 10.1080/17439760.2016.1262613]
37. Resnicow K, McMaster F, Bocian A, et al. Motivational interviewing and dietary counseling for obesity in primary care: an RCT. Pediatrics. Apr 2015;135(4):649-657. [doi: 10.1542/peds.2014-1880] [Medline: 25825539]
38. Schwartz RP, Hamre R, Dietz WH, et al. Office-based motivational interviewing to prevent childhood obesity: a feasibility study. Arch Pediatr Adolesc Med. May 2007;161(5):495-501. [doi: 10.1001/archpedi.161.5.495] [Medline: 17485627]
39. van Grieken A, Ezendam NPM, Paulis WD, van der Wouden JC, Raat H. Primary prevention of overweight in children and adolescents: a meta-analysis of the effectiveness of interventions aiming to decrease sedentary behaviour. Int J Behav Nutr Phys Act. May 28, 2012;9(1):61. [doi: 10.1186/1479-5868-9-61] [Medline: 22640437]
40. Campbell KJ, Crawford DA, Salmon J, Carver A, Garnett SP, Baur LA. Associations between the home food environment and obesity-promoting eating behaviors in adolescence. Obesity (Silver Spring). Mar 2007;15(3):719-730. [doi: 10.1038/oby.2007.553] [Medline: 17372323]
41. Birch LL, Ventura AK. Preventing childhood obesity: what works? Int J Obes (Lond). Apr 2009;33 Suppl 1:S74-S81. [doi: 10.1038/ijo.2009.22] [Medline: 19363514]
42. Redsell SA, Edmonds B, Swift JA, et al. Systematic review of randomised controlled trials of interventions that aim to reduce the risk, either directly or indirectly, of overweight and obesity in infancy and early childhood. Matern Child Nutr. Jan 2016;12(1):24-38. [doi: 10.1111/mcn.12184] [Medline: 25894857]
43. Direito A, Carraça E, Rawstorn J, Whittaker R, Maddison R. mHealth technologies to influence physical activity and sedentary behaviors: behavior change techniques, systematic review and meta-analysis of randomized controlled trials. Ann Behav Med. Apr 2017;51(2):226-239. [doi: 10.1007/s12160-016-9846-0] [Medline: 27757789]
44. Stephens J, Allen J. Mobile phone interventions to increase physical activity and reduce weight: a systematic review. J Cardiovasc Nurs. 2013;28(4):320-329. [doi: 10.1097/JCN.0b013e318250a3e7] [Medline: 22635061]

45. Hekler EB, Klasnja P, Riley WT, et al. Agile science: creating useful products for behavior change in the real world. Transl Behav Med. Jun 2016;6(2):317-328. [doi: 10.1007/s13142-016-0395-7] [Medline: 27357001]
46. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res. Jun 29, 2017;19(6):e232. [doi: 10.2196/jmir.7126] [Medline: 28663162]
47. Bickmore TW, Picard RW. Establishing and maintaining long-term human-computer relationships. ACM Trans Comput-Hum Interact. Jun 2005;12(2):293-327. [doi: 10.1145/1067860.1067867]
48. Miner AS, Shah N, Bullock KD, Arnow BA, Bailenson J, Hancock J. Key considerations for incorporating conversational AI in psychotherapy. Front Psychiatry. 2019;10:746. [doi: 10.3389/fpsyt.2019.00746] [Medline: 31681047]
49. Abd-Alrazaq AA, Rababeh A, Alajlani M, Bewick BM, Househ M. Effectiveness and safety of using chatbots to improve mental health: systematic review and meta-analysis. J Med Internet Res. Jul 13, 2020;22(7):e16021. [doi: 10.2196/16021] [Medline: 32673216]
50. Barlow SE, Expert Committee. Expert committee recommendations regarding the prevention, assessment, and treatment of child and adolescent overweight and obesity: summary report. Pediatrics. Dec 2007;120 Suppl 4:S164-S192. [doi: 10.1542/peds.2007-2329C] [Medline: 18055651]
51. Bower P, Gilbody S. Stepped care in psychological therapies: access, effectiveness and efficiency. Narrative literature review. Br J Psychiatry. Jan 2005;186:11-17. [doi: 10.1192/bjp.186.1.11] [Medline: 15630118]
52. Mohr DC, Burns MN, Schueller SM, Clarke G, Klinkman M. Behavioral intervention technologies: evidence review and recommendations for future research in mental health. Gen Hosp Psychiatry. 2013;35(4):332-338. [doi: 10.1016/j.genhosppsych.2013.03.008] [Medline: 23664503]
53. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res. Jan 30, 2015;17(1):e30. [doi: 10.2196/jmir.4055] [Medline: 25639757]
54. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med. Jun 2017;7(2):254-267. [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
55. Latulippe K, Hamel C, Giroux D. Social health inequalities and eHealth: a literature review with qualitative synthesis of theoretical and empirical studies. J Med Internet Res. Apr 27, 2017;19(4):e136. [doi: 10.2196/jmir.6731] [Medline: 28450271]
56. Eysenbach G. Design and evaluation of consumer health information web sites. In: Lewis D, Eysenbach G, Kukafka R, Stavri PZ, Jimison HB, editors. Consumer Health Informatics: Informing Consumers and Improving Health Care. Springer; 2005:34-60. [doi: 10.1007/0-387-27652-1_4]
57. Kohl LFM, Crutzen R, de Vries NK. Online prevention aimed at lifestyle behaviors: a systematic review of reviews. J Med Internet Res. Jul 16, 2013;15(7):e146. [doi: 10.2196/jmir.2665] [Medline: 23859884]
58. Følstad A, Brandtzaeg PB. Users' experiences with chatbots: findings from a questionnaire study. Qual User Exp. Dec 2020;5(1). [doi: 10.1007/s41233-020-00033-2]
59. Daniels SR, Hassink SG, Committee on Nutrition. The role of the pediatrician in primary prevention of obesity. Pediatrics. Jul 2015;136(1):e275-e292. [doi: 10.1542/peds.2015-1558] [Medline: 26122812]
60. Faith MS, Van Horn L, Appel LJ, et al. Evaluating parents and adult caregivers as "agents of change" for treating obese children: evidence for parent behavior change strategies and research gaps: a scientific statement from the American Heart Association. Circulation. Mar 6, 2012;125(9):1186-1207. [doi: 10.1161/CIR.0b013e31824607ee] [Medline: 22271754]

Abbreviations

ACT: Acceptance and Commitment Therapy
BAP: Brief Action Planning
CES-LCC: Comprehensive Evaluation Scale for LLM-Powered Counseling Chatbots
LLM: large language model
mHealth: mobile health
MI: motivational interviewing
ORBIT: Obesity-Related Behavioral Intervention Trials
SMART: Specific, Measurable, Achievable, Relevant, and Time-Bound
uMARS: user version of the Mobile Application Rating Scale

Edited by Matthew Balcarras; peer-reviewed by Alessandro Gaudino, Catarina Reis; submitted 19.Mar.2026; initial comments to author 09.Sep.2026; final revised version received 29.Sep.2026; accepted 14.Oct.2026; published 26.Nov.2026

Please cite as:
Purin B, Nicoletti AE, Bolpagni M, De Carli S, Devilli L, Rizzi S, Poggianella S, De Luca C, Stagi S, Franceschi R
Psychoeducational Chatbot (Demetra) for Preventing Childhood Overweight: Development and Usability Study
JMIR Pediatr Parent 2026;9:e95697
URL: https://pediatrics.jmir.org/2026/1/e95697
doi: 10.2196/95697

© Barbara Purin, Anna Elena Nicoletti, Marco Bolpagni, Simone De Carli, Lara Devilli, Silvia Rizzi, Stefania Poggianella, Chiara De Luca, Stefano Stagi, Roberto Franceschi. Originally published in JMIR Pediatrics and Parenting (https://pediatrics.jmir.org), 26.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Pediatrics and Parenting, is properly cited. The complete bibliographic information, a link to the original publication on https://pediatrics.jmir.org, as well as this copyright and license information must be included.