Intelligent Support for Digital Wellbeing: a Design Framework through a Systematic Literature Review

Int. J. Human–Computer Studies 205 (2025) 103653

Contents lists available at ScienceDirect

International Journal of Human - Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

# Intelligent support for digital wellbeing: A design framework through a systematic literature review ☆

Luca Scibetta *,[1], Massimiliano Pellegrino[1], Alberto Monge Roffarello, Luigi De Russis

*Department of Control and Computer Engineering, Politecnico di Torino, Torino, 10129, Italy*

ARTICLE INFO

*Keywords:*
AI
Digital wellbeing
Mental health
AI-applications for digital wellbeing
Systematic literature review

ABSTRACT

Recent advancements in AI, particularly Generative AI (GenAI) and Large Language Models (LLMs), have facilitated the integration of AI techniques into digital wellbeing applications, i.e., digital tools that aim at helping people's wellbeing as a sum of mental and emotional wellness. These AI-powered systems hold the potential to foster healthier habits by collecting and analyzing user behavioral data to provide personalized and dynamic solutions tailored to each user's needs and lifestyle, therefore improving the efficacy with respect to traditional non-AI interventions. Yet, their development presents significant challenges, including ethical concerns, privacy risks, and the potential for over-reliance on automated interventions. In this paper, we conduct a systematic literature review to examine the key characteristics, challenges, and opportunities in the existing research about AI-powered digital wellbeing tools. Based on our findings, we propose a design framework that outlines 6 critical dimensions and 23 sub-dimensions, spacing from user data and privacy to intervention strategies and personalization, offering practical guidance for researchers and practitioners developing AI-powered digital wellbeing applications. The framework emphasizes the importance of developing tailored and adaptive user-centered interventions adhering to scientific principles, psychological models and responsible data collection. We discuss the applicability and utility of our framework in evaluating and guiding the integration of AI in digital wellbeing applications.

## 1. Introduction

The World Health Organization (WHO) defines wellbeing a positive state experienced by individuals and societies, which is determined by social, economic, and environmental conditions. A specific aspect of wellbeing that has gained attention in recent years is *digital wellbeing*. However, there is currently no consensus on a universally accepted definition, and the term takes on various nuances. Some definitions focus on narrow aspects, such as associating digital wellbeing with satisfaction and time management in technology use (Zhao et al., 2024; Monge Roffarello and De Russis, 2019), or approaching it primarily through the lens of digital health and its implications on healthcare (Smits et al., 2022). Others offer broader perspectives: Floridi (Floridi, 2014) defines digital wellbeing as the influence of digital technologies on living a fulfilling and meaningful life within an information society. These definitions highlight many opportunities and challenges of technology's role in supporting people's wellbeing. Building on this literature, we adopt throughout this paper a targeted definition of digital wellbeing: we consider it as the contribution of digital technologies to users' mental and emotional wellbeing, either by providing tools that actively support psychological health or by trying to counteract negative consequences of technology use on people's emotional state. A wide range of applications fall under this definition, ranging from digital self-regulation tools (Roffarello and De Russis, 2023) to solutions that enhance empowerment and connectedness for neurotypical children (Stefanidi, 2022).

Recent advancements in Artificial Intelligence (AI), including, among the others, the recent diffusion of Generative AI (GenAI) and Large Language Models (LLMs), have further transformed the digital landscape, enabling the rapid proliferation of AI-powered applications (Bandi et al., 2023). These innovations have expanded into various domains, including healthcare (Rajashekar et al., 2024), education (Zhai et al., 2021), and, increasingly, the field of digital wellbeing.

AI-supported applications for digital wellbeing are now widespread, ranging from mental health chatbots (Abd-Alrazaq et al., 2021) to

☆ This article is part of a Special issue entitled: 'HCAI' published in International Journal of Human - Computer Studies.

* Corresponding author.
*E-mail addresses:* luca.scibetta@polito.it (L. Scibetta), massimiliano.pellegrino@polito.it (M. Pellegrino), alberto.monge@polito.it (A. Monge Roffarello), luigi.derussis@polito.it (L. De Russis).

[1] These authors contributed equally to the work.

https://doi.org/10.1016/j.ijhcs.2025.103653
Received 17 December 2024; Received in revised form 22 September 2025; Accepted 30 September 2025
Available online 8 October 2025
1071-5819/

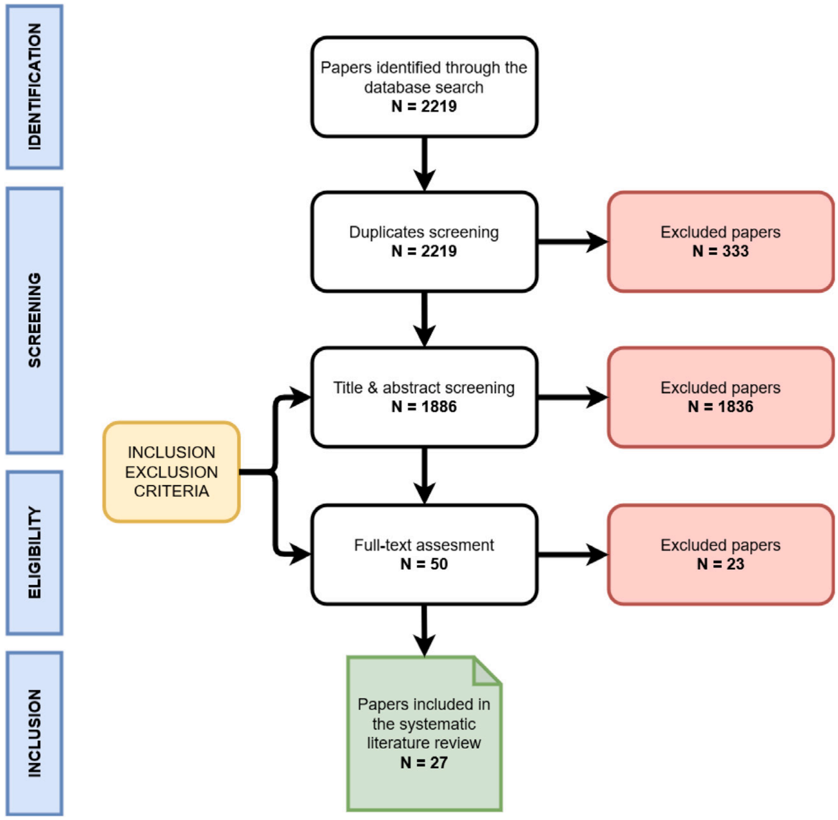



**Fig. 1.** Flowchart outlining the review process following the PRISMA guidelines.

mindfulness tools (Wang et al., 2024). However, despite their popularity, many of these applications are developed without adhering to robust scientific principles, comprehensive design guidelines or proper experimental validation (Casu et al., 2024). Given the sensitive nature of wellbeing and the potential risks associated with improper handling of user data or poorly designed interventions, there is a need to address this gap. Neglecting these considerations can inadvertently cause harm, both to individuals' mental and emotional health and to their privacy and trust in digital systems.

This study aims to bridge this gap by providing researchers and developers with a structured design framework for creating AI-integrated digital wellbeing applications. Following an approach similar to Victorelli et al. (2020), we conducted a systematic literature review to identify current trends, challenges, and opportunities in the design of AI-powered digital wellbeing applications and we used the results of the review as the basis to form our design framework, intended as a practical guide to practitioners. Our review primarily focuses on research works in the HCI community, as researchers in this field have explored digital wellbeing extensively (Monge Roffarello and De Russis, 2019; Lyngs et al., 2024; Parry et al., 2023).

Our goal is twofold: first, to advance future research by offering a comprehensive overview of the key factors and considerations in this field; and second, to protect and prioritize the wellbeing of users through ethically sound and scientifically validated design principles.

The systematic literature review and the design of the framework were guided by the following research questions:

- **RQ1**: What research objectives have driven existing studies on AI-powered digital wellbeing applications?
- **RQ2**: Which research methodologies are employed to integrate, utilize, and evaluate AI in digital wellbeing applications?
- **RQ3**: What challenges or potential issues arise from the use of AI in this domain, including ethical concerns, privacy risks, and unintended harm to users?
- **RQ4**: What design recommendations can be derived from existing studies for the development of AI-powered tools that support digital wellbeing?
- **RQ5**: How can the existing research on AI-powered digital wellbeing applications be synthesized into a design framework to guide future research?

Our findings form the foundation of the proposed design framework, which encompasses 6 critical dimensions and 23 sub-dimensions, spacing from user data and privacy to intervention strategies and personalization. Its aim is to provide a comprehensive understanding of the current state of the art and identify actionable insights for both researchers and developers. These insights are critical to foster the ethical and effective development of AI-integrated applications that prioritize user wellbeing, as well as to uncover gaps and opportunities for future advancements in this field.

While existing frameworks offer valuable guidance for integrating AI into various aspects of healthcare, considering for example clinical settings (Gama et al., 2022), healthcare practice (Lekadir et al., 2025), and specific innovations such as digital twins (Pellegrino et al., 2024); the framework we propose addresses a less-explored yet equally critical dimension: integrating AI into digital wellbeing applications. Our framework provide a structured, user-centered approach that supports the development of responsible, personalized, and adaptive AI-powered wellbeing tools. We have developed a wide and multi-sided guide accounting a number of different aspects to provide a single source of support. Thus, our work complements existing frameworks that focus on individual dimensions included in our framework, e.g., (Kip et al., 2025; Mummah et al., 2016; Michie et al., 2011).

This paper is structured as follows. Section 2 reports on our review and analysis methods. Section 3 discusses the results that emerged from the literature review. In Section 4, we present the design framework, that we built starting from the results of the review, with its 6 dimensions and 23 subdimensions. In Section 5, we discuss our results, explore how our framework can benefit researchers, and provide a detailed case study that demonstrates the utility of the design framework. Eventually, Section 6 summarizes the conclusions of our research.

## 2. Methodology

Our process for identifying and selecting relevant articles for this review was guided by the PRISMA framework for systematic literature reviews (Page et al., 2021). The procedural workflow is illustrated in Fig. 1.

### 2.1. Article selection

We conducted our literature search using the ACM Guide to the Computing Literature, the leading database in computing and human–computer interaction (HCI) research. This resource provided comprehensive access to full-text articles from conference proceedings, journals, books, and abstracts from prominent publishers, including ACM, IEEE, Springer, Elsevier, John Wiley & Sons, and Kluwer.

The final selection for this review was based on a search performed on October 3, 2024, focusing on articles published from September 1, 2019, to September 30, 2024. The search was further refined by selecting "research article" as the content type, excluding extended abstracts and short papers to focus on more comprehensive studies. In general, we prioritized research papers that address challenges, key considerations, and ethical aspects related to designing AI-powered tools for digital wellbeing. Keywords such as "AI", "LLM", and "chatbot" were combined with terms like "mental health" and "digital wellbeing" in our queries. We also included terms like "design framework", "mobile application", and "personalized intervention" to identify similar contributions that define design guidelines and to locate articles about tools that address digital wellbeing by tailoring individual user needs. The complete set of search queries is presented in Table 1. The initial search yielded 2,219 records, which were then reviewed by two authors to eliminate 333 duplicates and exclude 1,836 records that, upon examination of title and abstract, did not meet the screening criteria.

The selection criteria for inclusion focused on studies discussing digital tools (such as mobile, web, or browser applications) supporting wellbeing through the use of AI. We decided to account for all kinds of AI, for example also recommendation systems, but specifically looking for LLM and GenAI based systems. We also considered AI-powered digital wellbeing related chatbot. In the case of papers where the authors did not specify whether or not artificial intelligence was used in their work, given the ambiguity, we decided to include those whose results can also be applied to AI systems, such as those addressing personalization and customization. The examined tools could have been created by the study authors or examined through other means, such as literature reviews or analysis of user feedback.

The exclusion criteria were as follows:

- Studies centered on forms of digital wellbeing that fall outside of the definition we adopted, such as physical health (e.g. Bavaresco et al. (2024), Elnawawy et al. (2024)).
- Studies using AI for powering tangible and physical agents, such as robots or machines (e.g. Jingar and Lindgren (2019), Feijóo-García et al. (2024)).
- Studies focusing primarily on the technical aspects of AI/ML model development, without investigating their integration and user interaction in wellbeing tools (e.g. Gobin-Rahimbux et al. (2023), Chen et al. (2024)).

After a full-text review, we refined our selection, ultimately including 27 articles (16 conference papers and 11 journal articles) in our systematic review, hereafter referred to as the *General Corpus*. CHI[2] was the most frequently represented conference (10 articles), while the most common journal was PACMHCI[3] (2 articles). As shown in Fig. 2, our final corpus reveals that research about AI-powered digital wellbeing tools emerged within the last 5 years.

Within our *General Corpus*, we identified three categories based on each paper's main research contribution:

***Implemented Tools (N = 13):*** Studies that mainly focus on the implementation and, generally, user evaluation of a wellbeing supportive tool that integrates some type of AI, being it an existing tool (e.g. De Nieva et al. (2021)) or a novel developed one (e.g. Kim et al. (2024b)).

**Table 1**
The search queries used to search the electronic database of the ACM guide to the computing literature.

| Search query | # Results |
|---|---|
| ("LLM" OR "large language model" OR "AI" OR "artificial intelligence") AND ("guideline" OR "principle" OR "design space" OR "design framework") AND ("mental health" OR "mental wellbeing" OR "mental well-being" OR "mental wellness" OR "psychological health" OR "psychological wellbeing" OR "psychological well-being" OR "psychological wellness" OR "digital wellbeing" OR "digital well-being" OR "digital wellness" OR "digital health") | 1,341 |
| ("LLM" OR "large language model" OR "AI" OR "artificial intelligence") AND ("mobile application" OR "mobile app" OR "smartphone application" OR "smartphone app") AND ("mental health" OR "mental wellbeing" OR "mental well-being" OR "mental wellness" OR "psychological health" OR "psychological wellbeing" OR "psychological well-being" OR "psychological wellness" OR "digital wellbeing" OR "digital well-being" OR "digital wellness" OR "digital health") | 499 |
| ("LLM" OR "large language model" OR "AI" OR "artificial intelligence") AND ("web application" OR "web app" OR "webapp" OR "web platform") AND ("mental health" OR "mental wellbeing" OR "mental well-being" OR "mental wellness" OR "psychological health" OR "psychological wellbeing" OR "psychological well-being" OR "psychological wellness" OR "digital wellbeing" OR "digital well-being" OR "digital wellness" OR "digital health") | 180 |
| ("AI application" OR "AI app") AND ("mental health" OR "mental wellbeing" OR "mental well-being" OR "mental wellness" OR "psychological health" OR "psychological wellbeing" OR "psychological well-being" OR "psychological wellness" OR "digital wellbeing" OR "digital well-being" OR "digital wellness" OR "digital health") | 67 |
| ("LLM" OR "large language model") AND ("behavior change" OR "self control" OR "self regulation") | 65 |
| ("AI" OR "artificial intelligence" OR "LLM" OR "large language model") AND ("tailored solution" OR "personalized solution" OR "tailored intervention" OR "personalized intervention") AND ("mental health" OR "mental wellbeing" OR "mental well-being" OR "mental wellness" OR "psychological health" OR "psychological wellbeing" OR "psychological well-being" OR "psychological wellness" OR "digital wellbeing" OR "digital well-being" OR "digital wellness" OR "digital health") | 38 |
| ("chatbot" or "chat-bot" or "chat bot") AND ("mental health" OR "mental wellbeing" OR "mental well-being" OR "mental wellness" OR "psychological health" OR "psychological wellbeing" OR "psychological well-being" OR "psychological wellness" OR "digital wellbeing" OR "digital well-being" OR "digital wellness" OR "digital health") | 29 |

[2] https://dl.acm.org/conference/chi, last visited on November 12, 2024.

[3] https://dl.acm.org/journal/pacmhci, last visited on November 12, 2024.

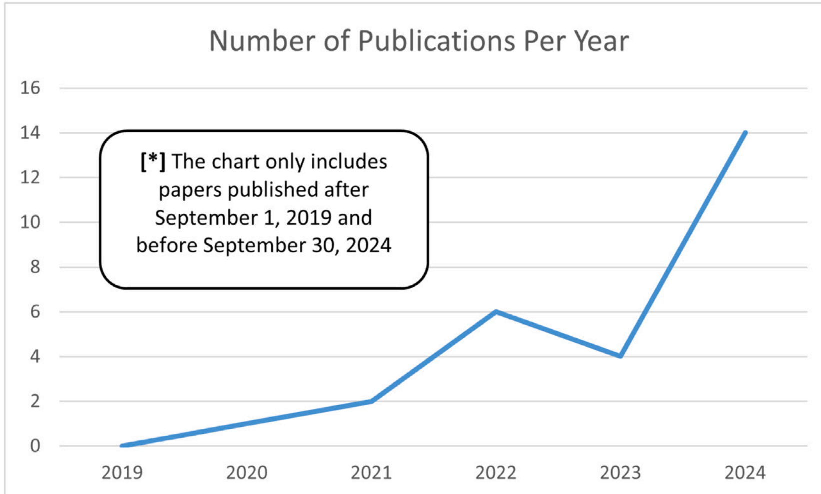



**Fig. 2.** Number of publications in our *General Corpus* per year, highlighting the rise of research on AI-powered wellbeing tools in the last 5 years.

***User Studies (N = 10):*** Studies regarding user surveys or interviews about wellbeing tools, without direct evaluation, such as experiences of use (Ma et al., 2024), participatory design (Feijóo-García et al., 2023), expert opinions (e.g., (Sweeney et al., 2021)), or evaluation of pre-generated model responses (Sharma et al., 2023).

***Comparative Works (N = 4):*** Studies comparing related works, including patent applications analyses (Karizat et al., 2024), user reviews of AI tools (Haque and Rubya, 2022), and two literature reviews (Valentine et al., 2022; Ahmad et al., 2022). We included the latter despite being literature reviews, focusing only on the authors' primary contributions in the form of considerations or design principles.

### *2.2. Data extraction*

To systematically extract data from our corpus, we developed a data extraction sheet that codes various elements relevant to our research questions. The initial columns of this sheet were used to characterize the articles under analysis, including the authors, title, abstract, publication type, and year of publication. We then extracted the *motivations* behind the discussion and analysis of AI solutions for digital wellbeing (**RQ1**). Additionally, we recorded data on the *methodologies* researchers used to approach these motivations, assess their work, and validate their results (**RQ2**). For each article, we also noted the *challenges* and ethical issues or implications encountered by researchers during their study (**RQ3**). We further extracted *design recommendations* regarding AI-powered digital wellbeing tools that emerged from each article (**RQ4**).

Moreover, we also applied three other coding schemas to better understand our corpus. The first two of these schemas were inspired by a related work on machine learning in mental health (Thieme et al., 2020). To begin with, we identified the main type of contribution presented in each article within our *General Corpus*, as previously discussed in Section 2.1 and shown in Fig. 3. Then we determined the specific wellbeing aspect targeted by the authors, whether it was *Mental Health* in general (e.g., (Haque and Rubya, 2022)), a more specific condition (e.g., *Affective Disorder* (Tong et al., 2023)), or a non-pathological wellbeing-related aspect (e.g., *Introspective Reflection* (Kim et al., 2024b; Seo et al., 2024)), as depicted in Fig. 4. Lastly, we coded how the interaction with the AI occurred in the tools implemented or studied. Often, it was in *Chatbot* modality (e.g., (De Nieva et al., 2021)), as visible in Fig. 5, but we also found tools based on more structured *Intervention Suggestions* (e.g., (Sharma et al., 2024)) or where the agent takes on the role of *Journaling Assistant* (e.g., (Kim et al., 2024b)). Three papers could not be integrated in this classification, as two of them (Capel et al., 2024; Karizat et al., 2024) are *Comparative Works* that do not focus on a specific *AI Interaction Type*, while another (Ahmad et al., 2022) explored some design principles for *Conversational Agents* in general, without focusing on a specific modality of conversation (chat or voice-based).

### *2.3. Framework development*

After analyzing the papers in our corpus, we summarized the finding into a design framework. We developed it through a mixed top-down and bottom-up approach, starting with our initial data extraction sheet where we annotated various noteworthy elements identified in the articles during the review phase. Through collaborative discussion among the authors, we grouped these elements and divided the articles into preliminary framework dimensions, establishing clear definitions for each dimension as the framework's first tier, representing the key areas to consider when designing AI-powered digital wellbeing tools.

Next, 5 articles out of the total 27 were randomly selected for another round of review, independently conducted by two authors. Each author independently analyzed how each article implemented, described, and evaluated AI-powered wellbeing tools, as well as their key findings, extracting additional noteworthy information and classifying it within the established dimensions while defining further subdimensions. After completing this independent analysis, the authors compared their work and collaboratively developed a provisional framework version, agreeing upon a coherent set of dimensions and subdimensions along with a clear analytical approach to ensure consistent classification of the remaining articles. This second tier details possibilities, options and considerations for researchers and practitioners during the design phase.

Finally, the same two authors independently analyzed and categorized the remaining 22 articles (11 each) using a shared classification sheet to categorize various aspects and characteristics of these articles into dimensions and subdimensions. Throughout this process, they continuously refined the framework structure as needed: creating new dimensions and subdimensions, modifying existing ones, merging or separating dimensions when appropriate, reflecting the bottom-up aspect of our approach. The authors regularly synchronized their work to ensure a coherent and consistently structured final framework, which will be described in detail in Section 4.

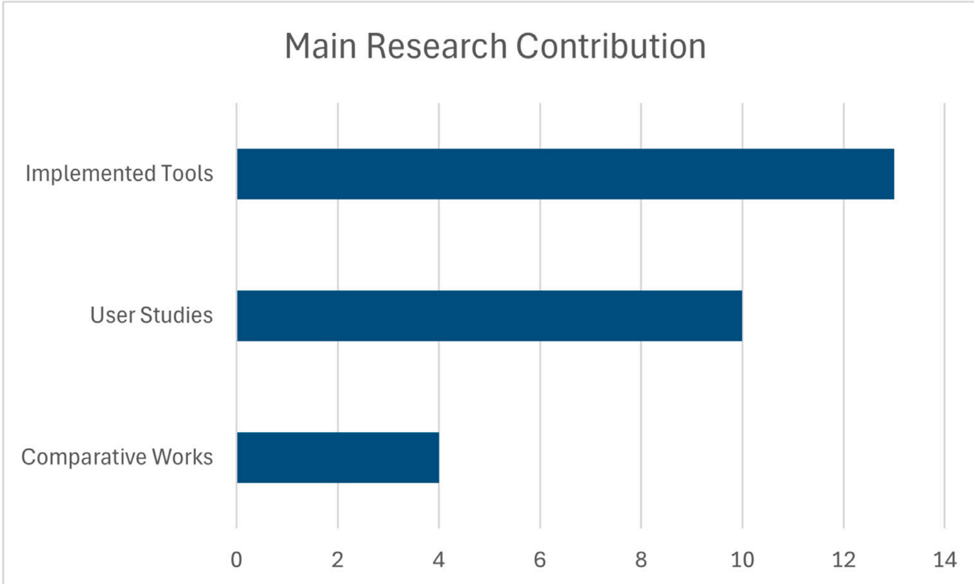



**Fig. 3.** Number of publications in our *General Corpus* per main research contribution, highlighting the prevalence of *Implemented Tools* and *User Studies* over *Comparative Works*.

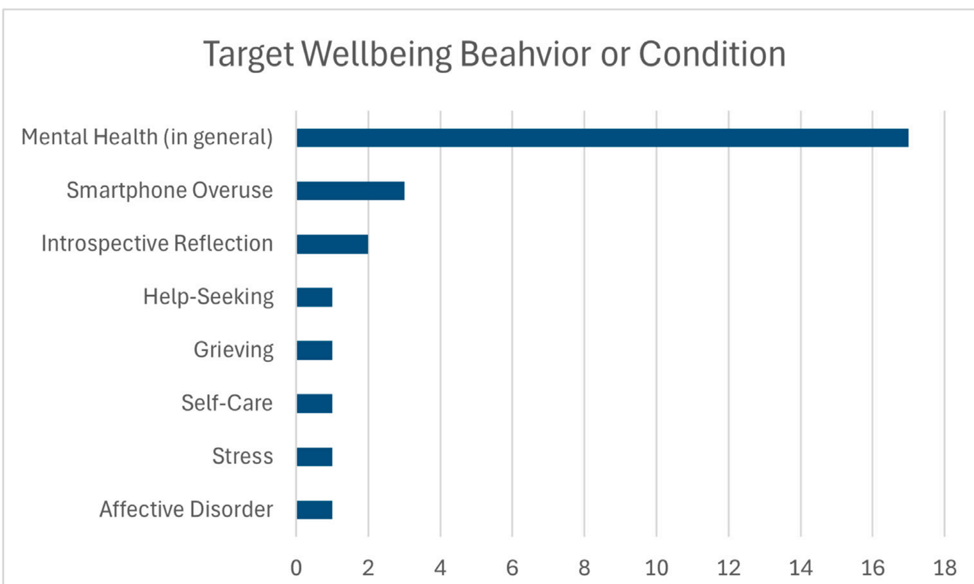



**Fig. 4.** Number of publications in our *General Corpus* per type of wellbeing behavior or condition that was the target of support, highlighting the strong prevalence of tools targeting *Mental Health* in general, over specific targeted conditions.

## 3. Results

In this section we report on the result of the analysis on our *General Corpus*, addressing each of our RQs in a different subsection.

### 3.1. Motivations

The review of our *General Corpus* revealed six primary motivations driving research in the domain of AI for digital wellbeing. Each category reflects distinct objectives and perspectives, summarized in Table 2.

The largest category (N = 13, (Sharma et al., 2023; Koulouri et al., 2022; Sharma et al., 2024; Monge Roffarello and De Russis, 2024; Yang et al., 2020; Valentine et al., 2022; Li et al., 2024; Capel et al., 2024; Cooney et al., 2024; Wu et al., 2024; Maharjan et al., 2022,?; Tong et al., 2023)) analyzes potential problems, challenges, possible solutions and future directions for integrating AI into digital wellbeing applications. These studies investigate the factors that influence acceptability, efficacy, supportiveness and user experience of AI wellbeing apps with respect to other solutions, also considering that it may be possible to adapt interventions and solutions over time to the user, eventually letting them regain their autonomy (Koulouri et al., 2022; Monge Roffarello and De Russis, 2024; Yang et al., 2020; Valentine et al., 2022; Li et al., 2024; Capel et al., 2024; Sharma et al., 2024; Wu et al., 2024; Tong et al., 2023). They also focus on promoting greater equity and access to treatment through AI applications, recognizing that therapy is not always affordable for everyone (Sharma et al., 2024; Valentine et al., 2022; Yang et al., 2020). In one case, for example, researchers saw the opportunity to add AI-powered interventions to a proper therapy to ease the work and improve effects (Cooney et al.,

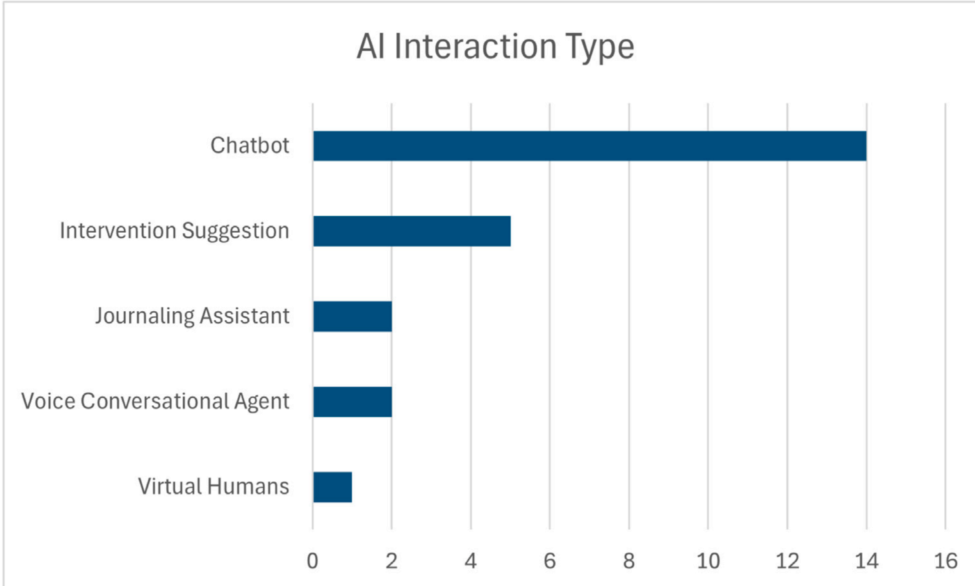



**Fig. 5.** Number of publications in our *General Corpus* per AI interaction type; it is notable how *Chatbots* are widely preferred.

**Table 2**
Summary table of motivations identified across the articles of our *General Corpus*.

| Motivation | Description |
|---|---|
| Changellenges & opportunities (N = 13) | Papers that uncover critical aspects and explore new possible directions for integrating AI into digital wellbeing applications. |
| Target user (N = 12) | Papers that reflect on the use of AI-powered digital wellbeing tools targeting specific users with specific needs. |
| User perception (N = 4) | How end users perceive AI-powered interventions for digital wellbeing. |
| Personalization (N = 3) | Papers that focus on making the proposed interventions personalizable to obtain better results. |
| Existing applications (N = 2) | Papers that focus on evaluating how existing applications work and how they are perceived. |
| Professional perception (N = 2) | How professionals perceive AI-powered interventions for mental wellbeing. |

2024). Two papers investigated problems regarding the interaction type, with reference to speech agents (Cooney et al., 2024; Maharjan et al., 2022). Other papers have examined the risks related to digital interventions, particularly regarding their invasiveness and the appropriateness of responses (Sharma et al., 2023; Capel et al., 2024; Yang et al., 2020).

The second category (N = 12, (Seo et al., 2024; Ma et al., 2024; Maharjan et al., 2022; De Nieva et al., 2021; Tong et al., 2023; Kim et al., 2024a; Feijóo-García et al., 2023; Xygkou et al., 2023; Young et al., 2024; Sharma et al., 2023; Koulouri et al., 2022; Yang et al., 2020)) focuses on the development of systems designed for specific types of user. These studies recognize the possible diversity of mental health needs across populations. While generalized solutions may fail to address the nuanced needs of different groups, potentially limiting their effectiveness, these articles aim to understand how incorporating demographic and psychological factors into AI digital wellbeing application design can produce solutions that are more accessible and effective. Emerging key factors include age (Seo et al., 2024; Young et al., 2024; Koulouri et al., 2022) and occupations, e.g., stressed students (De Nieva et al., 2021; Feijóo-García et al., 2023; Sharma et al., 2023; Yang et al., 2020) and workers (Tong et al., 2023). Ma et al. (2024) also highlight that sexual orientation has a significant impact on people's mental health, therefore suggesting to investigate how AI-powered tools for mental health are perceived by LGBTQ+ people (Ma et al., 2024). The analyzed papers suggest that a tailored solution should account for specific psychological conditions, like affective disorders (Maharjan et al., 2022), as well as, more broadly, the impact on psychiatric patients and their communication with mental health professionals (Kim et al., 2024a). Besides, people may sometimes want to use AI to manage temporary problems and needs, such as those experienced during grief (Xygkou et al., 2023).

The perception of end users is the focus of another four articles (N = 4, (Kim et al., 2024b; Young et al., 2024; Lee and Lee, 2024; Moilanen et al., 2022)), which aim at understanding how individuals interact with and perceive AI digital wellbeing applications. Underscoring the user perspective, these studies investigate the refining of this type of systems for a better alignment with expectations and needs of their audience, considering how talking and receiving suggestions from AI about sensitive and personal topics is perceived with its emotional response (Young et al., 2024; Kim et al., 2024b), how the graphical representation of the AI inside a chat influences its perception (Lee and Lee, 2024), and how users perceive chatbots with different personalities (Moilanen et al., 2022) or pretended demographics (Feijóo-García et al., 2023).

The fourth category (N = 3, (Ahmad et al., 2022; Wu et al., 2024; Vossen et al., 2024)) investigates the role of personalization. Building on the idea that AI can better comfort people and increase efficacy through personalization, these studies aim to understand how to optimize this personalization. Some researchers investigate how to design personality-adaptive conversational agents to improve interaction with users in mental health care (Ahmad et al., 2022). Others aim at understanding which type of personalization may be more desirable to users and how this influences therapeutic bind and usage intentions (Vossen et al., 2024), or how to personalize interventions and make them context aware with respect to momentarily mental states (Wu et al., 2024).

A further two articles (N = 2, (Haque and Rubya, 2022; Karizat et al., 2024)) take an evaluative approach by analyzing existing AI applications in the domain of digital wellbeing. While considering

**Table 3**
Summary table of methodological categories identified across the articles of our *General Corpus*.

| Methodological category | Description |
|---|---|
| Design and development (N = 16) | Papers that report on the design and development process of AI-powered tools or prototypes for digital wellbeing. |
| Testing (N = 15) | Papers that test AI tools or prototypes with users or experts. |
| Needfinding Interviews and surveys (N = 12) | Papers that conduct interviews and surveys with users and experts. |
| Reviews (N = 5) | Papers that report on literature reviews and analysis of users' reviews. |

**Table 4**
Challenges.

| Challenge category | Description |
|---|---|
| Ethical and safety concerns (N = 15) | Papers that discuss intrinsic biases, potential of inappropriate or harmful content generation, and risks of over-reliance on AI-systems. |
| Technical limitations and user experience (N = 14) | Papers that discuss context loss, emotional coldness, intervention inefficiencies, accessibility and evaluation limitations. |
| Privacy and data protection concerns (N=11) | Papers that discuss intrusive data requests, user worries over data management, and concerns about sharing intimate information with AI-systems. |

either AI patent application (Karizat et al., 2024) or user reviews from app stores (Haque and Rubya, 2022), these studies aim to identify the strengths and weaknesses of current implementations with a focus on usability, effectiveness, and ethical considerations.

Finally, two articles (N = 2, (Sweeney et al., 2021; Koulouri et al., 2022)) consider the perspective of mental healthcare professionals, exploring the implications of using AI-powered applications – particularly in the form of chatbots – in therapeutic contexts. These studies investigate how professionals perceive the potential of AI to complement traditional mental health services, reflecting on risks of misuse, such as over-reliance on these tools or unintended harm stemming from poorly designed interventions.

### 3.2. Methodologies

We identified four methodological categories employed in the research on AI applications for digital wellbeing, as summarized in Table 3.

The largest category (N = 16, (Koulouri et al., 2022; Seo et al., 2024; De Nieva et al., 2021; Vossen et al., 2024; Yang et al., 2020; Moilanen et al., 2022; Li et al., 2024; Monge Roffarello and De Russis, 2024; Sharma et al., 2024; Wu et al., 2024; Kim et al., 2024a,b; Maharjan et al., 2022; Tong et al., 2023; Feijóo-García et al., 2023; Capel et al., 2024)) includes the development of AI applications or prototypes for digital wellbeing. Most researchers decided to work on the design of chatbots (Koulouri et al., 2022; Seo et al., 2024; De Nieva et al., 2021; Vossen et al., 2024; Yang et al., 2020; Moilanen et al., 2022). We also found a paper describing a speech-enabled conversational assistant (Maharjan et al., 2022), and a subset of other apps that focused on supporting users' mental health, stress management, and healthy use of technology, through alternative methods like AI-enhanced journaling (Kim et al., 2024b,a) and recommender systems (Valentine et al., 2022; Tong et al., 2023). Interestingly, in some cases, researchers decided to proceed through participatory design sessions or workshops (Feijóo-García et al., 2023; Capel et al., 2024).

The second largest category (N = 15, (Sharma et al., 2024; De Nieva et al., 2021; Seo et al., 2024; Sharma et al., 2023; Koulouri et al., 2022; Kim et al., 2024b; Maharjan et al., 2022; Monge Roffarello and De Russis, 2024; Kim et al., 2024a; Tong et al., 2023; Wu et al., 2024; Li et al., 2024; Lee and Lee, 2024; Vossen et al., 2024; Moilanen et al., 2022)), which also includes some articles of the previous one, comprises direct experimentation with users or experts. Some researchers involved various and diverse users for testing (Sharma et al., 2024; De Nieva et al., 2021), while others focused on specific populations, e.g., children (Seo et al., 2024). Some involved only experts (Koulouri et al., 2022) or both users and experts (Sharma et al., 2023). Most of these studies were conducted on the field (Kim et al., 2024b; Maharjan et al., 2022; Monge Roffarello and De Russis, 2024; Kim et al., 2024a), with a duration lasting from a few days (Kim et al., 2024b) to a few weeks (Maharjan et al., 2022). Given the sensitive nature of the involved populations, other experiments were instead conducted in the lab (Seo et al., 2024). Many researchers decided to carry out between-subjects tests (Tong et al., 2023; Wu et al., 2024; Li et al., 2024; Lee and Lee, 2024; Vossen et al., 2024), with only one study applying a within-subjects methodology (Moilanen et al., 2022).

Data collection through interviews and surveys is another prominent category (N = 12, (De Nieva et al., 2021; Ahmad et al., 2022; Young et al., 2024; Koulouri et al., 2022; Sweeney et al., 2021; Seo et al., 2024; Ma et al., 2024; Capel et al., 2024; Maharjan et al., 2022; Xygkou et al., 2023; Cooney et al., 2024; Vossen et al., 2024)). These are studies in which authors may or may not have designed, developed, or tested AI implementations, but either way have included a focus on user and expert feedback, exploring themes such as user experience, ethical concerns, and perceived impact. Interviews were aimed either at experts (Seo et al., 2024; Cooney et al., 2024), users (Ma et al., 2024; Capel et al., 2024), or particular categories of people, e.g., people coping with grief (Xygkou et al., 2023), children (Seo et al., 2024). In one case (Maharjan et al., 2022), both interviews and a post-study survey were carried out. In many other cases, only surveys were conducted, often online (Sweeney et al., 2021). Surveys, like interviews, were directed at users (Young et al., 2024; Ahmad et al., 2022) or experts (Sweeney et al., 2021), and included a mix of pre-study and post-study deliveries (Xygkou et al., 2023). Moreover, it happened that experts were interpellated directly to give their opinion and provide a form of expert validation of tools or design principles (Vossen et al., 2024; Ahmad et al., 2022).

The fourth methodological category (N = 5, (Haque and Rubya, 2022; Karizat et al., 2024; Koulouri et al., 2022; Ahmad et al., 2022; Valentine et al., 2022) comprises papers that use a review as a mean of research. While some researchers took out a literature review (Koulouri et al., 2022; Ahmad et al., 2022; Valentine et al., 2022), others focused on gray literature (Karizat et al., 2024) or other sources, like users' reviews on app stores (Haque and Rubya, 2022).

### 3.3. Challenges

Our review reveals three main categories of challenges and issues that researchers have consistently documented across studies and implementations, summarized in Table 4.

The first challenge refers to ethical and safety concerns surrounding the use of AI for digital wellbeing (N = 15 (Xygkou et al., 2023; Wu et al., 2024; Sharma et al., 2024; Sweeney et al., 2021; Vossen et al., 2024; Karizat et al., 2024; Ma et al., 2024; Sharma et al., 2023; Valentine et al., 2022; Seo et al., 2024; Cooney et al., 2024; Young et al., 2024; Kim et al., 2024b,a; Capel et al., 2024)). Xygkou et al. (2023), for example, questioned the appropriateness of simulating conversations with deceased individuals to cope with grief (Xygkou et al., 2023). Other researchers addressed potential inequities in responses and performances among specific user subpopulation caused by LLMs' biases (Sharma et al., 2024). For example, the work by Ma et al. (2024) shows that LGBTQ+ people are sometimes stereotypically considered as a monolithic group, neglecting the specificity of needs and issues of each person in the community and not keeping in pace with evolving social norms. Researchers also highlight the risk of generating harmful or unpleasant content in prompt responses (Wu et al., 2024; Sharma et al., 2024; Cooney et al., 2024; Young et al., 2024; Seo et al., 2024; Kim et al., 2024b), as well as false positive recommendations, particularly for users with severe pathologies (Valentine et al., 2022). Safety concerns arise regarding suicide prevention and crisis response adequacy (Sweeney et al., 2021; Vossen et al., 2024; Kim et al., 2024a). A study that deals with mental health emphasize avoiding stigmatization and criminalization of behaviors when leveraging AI for digital wellbeing (Karizat et al., 2024). Other researchers warn that overly long or dense responses can be counterproductive and harmful to users seeking support (Sharma et al., 2023), and that chatbots designed to encourage emotional openness may result in opposite effects (Seo et al., 2024). Authors note risks of system dependency (Sharma et al., 2024), emotional attachment to technology (Capel et al., 2024), and users failing to develop the ability to express their own emotions, instead relying on AI-generated expressions (Kim et al., 2024b). There is also a consistent warn against over-reliance on AI as therapeutic replacements (Ma et al., 2024; Young et al., 2024; Kim et al., 2024b; Cooney et al., 2024). In a study testing LLM-supported journaling for psychiatric patients, it was found that some people exaggerated their conditions when interacting with the system (Kim et al., 2024a).

Our *General Corpus* also reveals significant technical constraints affecting user experience (N = 14 (Haque and Rubya, 2022; Ma et al., 2024; Sweeney et al., 2021; Cooney et al., 2024; Maharjan et al., 2022; Seo et al., 2024; Vossen et al., 2024; Monge Roffarello and De Russis, 2024; De Nieva et al., 2021; Moilanen et al., 2022; Li et al., 2024; Kim et al., 2024a; Tong et al., 2023; Valentine et al., 2022)), with chatbots that frequently lose conversation context (Haque and Rubya, 2022), deviate from given instructions (Seo et al., 2024), and provide generic (Ma et al., 2024), repetitive (Haque and Rubya, 2022), and even uncertain or inaccurate (Kim et al., 2024a) responses. In addition, two studies targeting AI-powered chatbots demonstrated limited emotional understanding by these tools (Sweeney et al., 2021; Haque and Rubya, 2022), and other two studies showed how frequent communication breakdowns create frustration (De Nieva et al., 2021), especially in sensitive moments when the chatbot is really needed (Xygkou et al., 2023). Similarly, evidence shows that children sometimes find chatbots uninteresting over time (Seo et al., 2024), and that chatbots are perceived as lacking warmth, regardless of the personality assigned to them (Moilanen et al., 2022). Regarding recommender systems, a study about working from home related stress found that AI-based recommendations are not associated to significant improvement (Tong et al., 2023), and another discussed the risks of choice overload when users are provided with too many recommendations (Valentine et al., 2022). Two works expressed limitations in smartphone overuse interventions, as repetitive prompts to stop using apps can even worsen it (Li et al., 2024) and systems are not capable to always distinguish compulsive from normal phone use (Li et al., 2024). We also found accessibility issues, as voice interaction proves problematic without private spaces (Cooney et al., 2024; Maharjan et al., 2022). Moreover, conversational systems are not capable to support people that struggle to describe emotions due to apathy or cognitive load, according to Kim et al. (2024a), and in general digital wellbeing systems require for people to own a personal device, which is not always the case, especially when children are the target of intervention (Seo et al., 2024). A limitation identified in two studies concerning experimental evaluations is the requirement for prolonged observation periods to ensure reliable results, especially when examining behavioral modifications (Vossen et al., 2024; Monge Roffarello and De Russis, 2024).

**Table 5**
Design recommendations.

| Recommendation | Description |
|---|---|
| Personalization and adaptation (N = 16) | Use user data, preferences, and context to provide tailored support. |
| Professional support (N = 15) | Ground interventions in psychological techniques, complementing mental health specialists and providing adequate crisis response. |
| Transparent data Use (N = 14) | Communicate clearly about AI, data handling, and privacy, while safeguarding sensitive information. |
| Emotional connection (N = 10) | Create a safe, empathetic, and engaging space where users feel supported and understood. |
| Reflective and goal orientated interventions (N = 8) | Encourage introspection, critical thinking, and autonomy through goals and gradual support reduction. |

Finally, the analyzed studies highlight privacy and data protection as a major concern in AI-powered digital wellbeing applications (N = 11 (Lee and Lee, 2024; Maharjan et al., 2022; Yang et al., 2020; Cooney et al., 2024; Kim et al., 2024a; Wu et al., 2024; Capel et al., 2024; Sharma et al., 2023; Sweeney et al., 2021; Xygkou et al., 2023; Koulouri et al., 2022)), emphasizing careful handling of data belonging to sensitive user populations (Cooney et al., 2024; Kim et al., 2024a; Sweeney et al., 2021; Koulouri et al., 2022) and people close to them (Xygkou et al., 2023), especially when data is shared with third-party (Wu et al., 2024; Yang et al., 2020). Indeed, AI-based systems often leverage private cloud-based models. Consequently, users often express anxiety about sharing personal, intimate and emotional information with conversational agents (Capel et al., 2024). According to Sharma et al. (2023), in particular, an AI-powered digital wellbeing tool that request more data than necessary becomes intrusive. Finally, one article highlights how privacy worries amplifies when chatbots are perceived as having thoughts and emotions, therefore similar to humans who could judge users (Lee and Lee, 2024). Privacy concerns also extend to voice interactions, as users are worried about constant listening (Maharjan et al., 2022).

### 3.4. Design recommendations

Finally, we synthesize the design recommendations for AI-powered digital wellbeing tools that arise from our *General Corpus*. As reported in Table 5, five main recommendations areas emerge throughout the analyzed papers.

Our analysis reveals that personalization and adaptation are crucial design elements (N = 16 (Moilanen et al., 2022; Ahmad et al., 2022; Vossen et al., 2024; Valentine et al., 2022; Capel et al., 2024; Wu et al., 2024; Feijóo-García et al., 2023; Seo et al., 2024; Haque and Rubya, 2022; Ma et al., 2024; Maharjan et al., 2022; Monge Roffarello and De Russis, 2024; Tong et al., 2023; Valentine et al., 2022; Lee and Lee, 2024; Cooney et al., 2024)). Notably, research identifies current limitations in existing chatbot systems, including overly generic responses that do not take into consideration the individuality of the user (Ma et al., 2024) and insufficient personalization of solutions (Haque and Rubya, 2022). One study highlights that AI-driven adaptive interventions outperform randomized approaches (Tong et al., 2023). Systems, therefore, could personalize the interventions by appropriately asking for user data (Wu et al., 2024) and consequently adapt to their momentary needs and contextual factors (Moilanen et al., 2022). This can be made via automatic detection or provision of multiple options. Smartphone data can be used to understand behavioral patterns – from physical inactivity (Valentine et al., 2022) to mental wellbeing issues and smartphone overuse (Monge Roffarello and De Russis, 2024) – and consequently recommend interventions.

Multiple dimensions can be customizable, including the communication style, the role taken by the agent (Ahmad et al., 2022) and its therapeutic approach (Vossen et al., 2024; Valentine et al., 2022), and the interaction modalities (Capel et al., 2024; Maharjan et al., 2022). Demographic aspects of the agent could also be made customizable, taking in consideration that users seem to express preference when agents are similar to themselves (Feijóo-García et al., 2023), and that one article shows how anthropomorphic profile images enhance engagement (Lee and Lee, 2024). Despite this, careful balance is needed between providing sufficient options or recommendations and avoiding choice overload (Valentine et al., 2022), taking into account the specific needs of different user populations (Seo et al., 2024). Finally, it is also emphasized the importance of filtering the truly relevant information from user-shared data to deliver an efficient targeted support (Cooney et al., 2024).

The analyzed articles also highlight the importance of grounding tools in established psychological approaches to deliver a professional support while maintaining appropriate boundaries with the actual therapist (N = 15 (Sweeney et al., 2021; Capel et al., 2024; Koulouri et al., 2022; De Nieva et al., 2021; Kim et al., 2024a; Ahmad et al., 2022; Haque and Rubya, 2022; Moilanen et al., 2022; Kim et al., 2024b; Ma et al., 2024; Valentine et al., 2022; Xygkou et al., 2023; Young et al., 2024; Cooney et al., 2024; Vossen et al., 2024)). Some authors suggest to implement evidence-based techniques like Cognitive Behavioral Therapy (CBT) (Haque and Rubya, 2022; De Nieva et al., 2021) and Self Determination Theory (SDT) (Valentine et al., 2022). Other indicates that involving mental health experts in design (Koulouri et al., 2022; Kim et al., 2024a) and showing competence and professionalism (Ahmad et al., 2022; Moilanen et al., 2022; Tong et al., 2023; Capel et al., 2024) enhances tool credibility, creating what one study (Valentine et al., 2022) calls “Digital Therapeutic Alliance”. Despite this, our corpus often emphasizes that AI should complement rather than replace mental health professionals (Young et al., 2024; Cooney et al., 2024), while supporting treatment through functions like therapeutic journaling (Kim et al., 2024a) and medication adherence reminders (Sweeney et al., 2021). In fact, people typically prefer relying on professionals for real treatments (Xygkou et al., 2023), with some articles highlighting concerns about users potentially over-relying on AI as therapeutic replacements (Ma et al., 2024; Kim et al., 2024b) and evidence showing that this over-reliance can also affect the relationship between the user and the human therapist when they exists (Kim et al., 2024a). Finally, AI should also be able to appropriately handle crisis situations (Sweeney et al., 2021), stepping aside and automatically notifying psychiatrists for self-harm or suicide risks (Kim et al., 2024a) or connecting directly to other emergency services and helplines (Haque and Rubya, 2022; Vossen et al., 2024; Ma et al., 2024; Sweeney et al., 2021).

As the third design recommendation, the literature emphasizes the importance of transparent communication about AI and data handling (N = 14 (Haque and Rubya, 2022; Ahmad et al., 2022; Sharma et al., 2023; Sweeney et al., 2021; Koulouri et al., 2022; Kim et al., 2024b; Ma et al., 2024; Yang et al., 2020; Wu et al., 2024; Valentine et al., 2022; Capel et al., 2024; Xygkou et al., 2023; Young et al., 2024; Cooney et al., 2024)), highlighting the critical importance of protecting sensitive user information, particularly in the mental health field (Cooney et al., 2024; Koulouri et al., 2022; Sweeney et al., 2021). Some articles express users’ concerns about privacy when sharing intimate personal information (Ma et al., 2024; Capel et al., 2024) and worries about who has access to the conversations and AI agents that continuously monitor them (Maharjan et al., 2022). Therefore, systems should clearly communicate AI involvement (Young et al., 2024; Haque and Rubya, 2022), data management practices (Ahmad et al., 2022), and system limitations (Kim et al., 2024b). Some technical solutions are documented, including encrypted conversations on dedicated servers (Yang et al., 2020) and open-source AI implementations for autonomous data handling (Wu et al., 2024). Some articles reveal that users perceive excessive data collection as intrusive (Sharma et al., 2023), therefore emphasizing the need for systems that propose personalized interventions to balance quality and quantity of collected data with recommendation accuracy (Valentine et al., 2022). Our corpus also addresses ethical considerations regarding data sharing, including third parties data management (Xygkou et al., 2023) and autonomous governance on data history, assuring transparency on how and to what extent it affects interventions (Valentine et al., 2022).

Our review also highlights perspectives on designing intelligent tools for digital wellbeing that are able to establish an emotional connection with users (N = 10 (Ahmad et al., 2022; Ma et al., 2024; Maharjan et al., 2022; Xygkou et al., 2023; Kim et al., 2024b; De Nieva et al., 2021; Moilanen et al., 2022; Karizat et al., 2024; Capel et al., 2024; Lee and Lee, 2024)). One article notes that some existing chatbots are perceived as emotionally distant (Moilanen et al., 2022), and generally it is emphasized the importance of balancing emotional engagement with appropriate technological boundaries. An article expressed how emulating capability of comprehension and expression of human-like emotions is important (Sweeney et al., 2021), and according to another study people also find it therapeutic to sometimes suspend incredulity and pretend to talk to a real person, a dear one (Xygkou et al., 2023). Two papers also demonstrate users value constant availability and proactive engagement (Ahmad et al., 2022; Haque and Rubya, 2022), while others place emphasis on creating a safe, non-judgmental environment (Ma et al., 2024; Xygkou et al., 2023). One study indicates that the agent’s capability of creating an emotional connection with the user outweighs technical efficiency in importance (Xygkou et al., 2023), with people appreciating dynamic, engaging conversations over robotic interactions, according to another article (Maharjan et al., 2022). Some studies show these tools should focus on making users feel understood and listened to De Nieva et al. (2021), Ahmad et al. (2022), Haque and Rubya (2022), also making them comfortable in talking (Karizat et al., 2024) and opening up (Kim et al., 2024b), rather than merely providing advice. However, some works also identifies potential risks, including emotional attachment to AI (Capel et al., 2024), resulting frustration when it fails to understand Xygkou et al. (2023), and increased privacy concerns when chatbots appear too human-like (Lee and Lee, 2024).

The last recommendation highlights the potential of AI systems to support reflective and goal oriented interventions (N = 8 (Xygkou et al., 2023; Kim et al., 2024b; Yang et al., 2020; Kim et al., 2024a; Seo et al., 2024; Sharma et al., 2024; Monge Roffarello and De Russis, 2024; Valentine et al., 2022)). Our review shows that these systems can facilitate critical self-examination through varied suggestions (Kim et al., 2024b) and multiple perspectives to overcome rigid viewpoints (Kim et al., 2024a), and supportive interventions that help to understand and address the root of the problems (Yang et al., 2020), with intervention proposals serving as reflection prompts even when not directly implemented (Valentine et al., 2022). A study emphasizes gradual reduction of support to develop user independence (Monge Roffarello and De Russis, 2024), and another similarly place importance on verifying users’ ability to maintain learned reasoning patterns autonomously (Sharma et al., 2024). A work highlights the value of clear goal-setting and task definition (Valentine et al., 2022), and others define specific goals AI tools can help to achieve, such as promoting social reintegration (Xygkou et al., 2023) and facilitating parent–child emotional discussions (Seo et al., 2024).

## 4. Framework

Our review has shed light on the motivations and methods that shape the development of AI-powered digital wellbeing tools, as well as on the challenges and recommendations researchers encounter or formulate along the way. We discovered that researchers typically motivate their works either with the exploration of challenges and opportunities in the field, such as acceptability or equity, or with the development and study of systems that target a specific user populations with specific needs, like psychiatric patients, stressed workers, or LGBTQ+ people. We also found that among the most common methodological approaches, there were studies that tested existing AI implementations, studies that designed and developed novel applications or prototypes, and others that conducted surveys or interviews with both general users and experts. Typical challenges encountered by authors include ethical and safety concerns, such as biased or inappropriate outputs; technical limitations, usually expressed as inefficient or repetitive outputs; and privacy concerns, typically represented by user worries about who has access to their intimate information. When reporting on design recommendations, we found that many authors emphasized the importance of having systems that adapt interventions to the specific user profile and needs. They also noted that people appreciate AI systems that are perceived as professional, transparent about their data management, and able to emotionally connect with them, orienting the intervention on self-reflection and clear goal definitions.

Overall, our findings highlight critical areas for improving the design of digital wellbeing tools based on AI to ensure safe, efficient, and engaging AI outputs. Furthermore, digital wellbeing AI systems, especially commercial ones, often lack scientific validation and a rigorous intervention approach (Haque and Rubya, 2022; Valentine et al., 2022). To close these gaps, this section proposes a comprehensive design framework grounded on our findings. As shown in Table 6, the framework is divided into two hierarchical dimensions. The first tier includes dimensions that represent the major areas to consider during the design of digital wellbeing tools that leverage AI, namely User Information, Intervention, Interaction, Data Management, Model, and Study. The second tier includes sub-dimensions that detail the parent dimension into more specific design aspects, such as Demographics and Lifestyle for the User Information, Purpose and Approach for the Intervention dimension, Role and Tone for the Interaction dimension, Transparency and Privacy for the Data Management dimension, Selection and Response Quality for the Model dimension, and Type and Participants for the Study dimension.

Below we present dimensions and sub-dimensions one by one, highlighting the bibliographic references each of the framework items is grounded on.

### 4.1. User information

The User Information dimension describes the user-related data that would be valuable for AI-powered digital wellbeing systems to collect and use in order to tailor the wellbeing intervention to the person who is using the tool.

#### 4.1.1. Demographics

Demographics describes user identity characteristics such as age and gender, as well as the wellbeing aspects the user needs to address. Related to age, systems can either target a specific age range (Koulouri et al., 2022; Seo et al., 2024) or personalize interventions based on the user's age. Sexuality and gender identity can be considered as well, especially if the designed tool aims to address issues such discrimination and marginalization of vulnerable populations, e.g., LGBTQ+ people (Ma et al., 2024). In this context, practitioners should be aware and address the inherent biases and stereotypes AI models often have (Lee et al., 2024; Gross, 2023). Furthermore, it is particularly important to consider collecting user's wellbeing status, which is the main reason that brings a person to seek support through digital support. This status may include pathologic disorders, such as affective disorder (Maharjan et al., 2022); a behavior condition, such as smartphone overuse (Wu et al., 2024); or a momentary unpleasant situation, such as work-related stress (Tong et al., 2023).

#### 4.1.2. Lifestyle

Besides demographics, AI-powered digital wellbeing tools can also take into account users lifestyle data. Occupations, hobbies, and users' habits (Wu et al., 2024; Tong et al., 2023), for example, can be useful to personalize interventions based on users' interests and the various contexts they encounter during the day. Indeed, our findings show that lifestyle-related information such as the occupation is linked to wellbeing struggle. For example, students and academics frequently manifest study-related issues like procrastination, lack of concentration, and difficulties in learning certain subjects (Yang et al., 2020; De Nieva et al., 2021).

#### 4.1.3. Goals

People's goals and objectives, particularly those related to wellbeing, are of major importance when it comes to develop digital wellbeing tools based on AI (Wu et al., 2024; Valentine et al., 2022). Taking into account this information is fundamental to address the specific and evolving user's needs, and people who are not able to find or express their goals should be helped by the tool to define their needs as a first step towards mitigating their wellbeing issues.

#### 4.1.4. Personal device metrics

Personal device metrics can offer key insights into users' wellbeing conditions and long term progress. For example, apps or websites usage data (Monge Roffarello and De Russis, 2024; Tong et al., 2023), as well as preferences around distractive apps or websites (Wu et al., 2024), can help the system to understand when the user is experiencing device overuse and intervene appropriately. Other metrics can be derived from the phone accelerometer and the GPS localization tracking (Valentine et al., 2022). The first can be used to understand if the user is being sedentary for a long time, a condition that is sometimes associated with mental health issues (Hoare et al., 2016; Huang et al., 2020), while the second can be exploited to deliver appropriate support based on the current user's context, such as distinguishing between private and public spaces or between home and work environments.

### 4.2. Intervention

The Intervention dimension offers insights on how to design AI-generated wellbeing paths that always deliver appropriate, harmless, and emergency-ready outputs, while taking into account the approach to be used and the purpose to be pursued during such paths.

#### 4.2.1. Purpose

Defining the purpose of the offered interventions is essential for them to be really effective, and that purpose should be clearly communicated to the user. Interventions can address the root causes of wellbeing issues, or offer a more immediate relief (Yang et al., 2020). AI interventions can be designed to encourage users to express their emotions with others (Seo et al., 2024), or they can aim to make the tool itself serve as the user's listener and advisor (Kim et al., 2024b). This approach can be particularly useful when the user is not yet ready or is not able to open up to others due to the sensitivity of topics, a lack of available listeners, or any other reason. Generally speaking, though, digital wellbeing support should have an "expiring date", not as a strict deadline, but rather as a tendency to offer users the instruments to become independent or able to seek human-help support (Xygkou et al., 2023; Monge Roffarello and De Russis, 2024; Sharma et al., 2024). Overall, it should be clear that the purpose of digital wellbeing tools is to *support* conventional professional therapy, and that experts and health professionals should be involved in the definition of methods and boundaries of such a digital support (Kim et al., 2024b).

#### 4.2.2. Approach

Choosing the right approach is also essential when designing an intervention. Consulting experts and practitioners to ground AI-based

**Table 6**
The design framework, divided into dimensions and sub-dimensions, each with its description and bibliographic references.

| Dimension | Sub-dimension | Description | References |
|---|---|---|---|
| User information | Demographics | Personal information about the user, including their health/wellbeing status. | N=6 Koulouri et al. (2022), Seo et al. (2024), Maharjan et al. (2022), Wu et al. (2024), Ma et al. (2024), Tong et al. (2023) |
| | Lifestyle | Data about daily life of the user concerning both work and free time. | N=4 Wu et al. (2024), Tong et al. (2023), Yang et al. (2020), De Nieva et al. (2021) |
| | Goals | Aspect(s) of their life that the user wants to improve and why. | N=2 Wu et al. (2024), Valentine et al. (2022) |
| | Personal Device Metrics | Data gathered from devices that provide insights on the user conditions. | N=4 Monge Roffarello and De Russis (2024), Wu et al. (2024), Valentine et al. (2022), Tong et al. (2023) |
| Intervention | Purpose | The objective that the proposed intervention should achieve, both in the moment and over time. | N=7 Seo et al. (2024), Kim et al. (2024a,b), Sharma et al. (2024), Monge Roffarello and De Russis (2024), Yang et al. (2020), Xygkou et al. (2023) |
| | Approach | The appropriate means that will allow reaching the identified purpose. | N=11 Haque and Rubya (2022), De Nieva et al. (2021), Kim et al. (2024b,a), Ahmad et al. (2022), Sharma et al. (2024), Tong et al. (2023), Valentine et al. (2022), Xygkou et al. (2023), Seo et al. (2024), Koulouri et al. (2022) |
| | Adaptation | How momentary situations, topics, and users' specific needs are recognized and managed. | N=6 Moilanen et al. (2022), Kim et al. (2024a), Young et al. (2024), Tong et al. (2023), Valentine et al. (2022), Li et al. (2024) |
| | Non-Maleficence | The precautions to apply to avoid direct harm to the user. | N=10 Wu et al. (2024), Sharma et al. (2023), Karizat et al. (2024), Cooney et al. (2024), Kim et al. (2024b), Young et al. (2024), Koulouri et al. (2022), Ma et al. (2024), Valentine et al. (2022), Seo et al. (2024) |
| | Crisis Management | The detection of emergency and harmful situations and the delivery of an appropriate response. | N=6 Haque and Rubya (2022), Sweeney et al. (2021), Kim et al. (2024a), Vossen et al. (2024), Seo et al. (2024), Ma et al. (2024) |
| Interaction | Role | Which social role the agent should assume when interacting with the user, optionally considering preferences. | N=7 Kim et al. (2024a), Capel et al. (2024), Xygkou et al. (2023), Young et al. (2024), Ahmad et al. (2022), Vossen et al. (2024), Cooney et al. (2024) |
| | Tone | With which degree of formality and confidence the agent should interact with the user, optionally considering preferences. | N=9 Haque and Rubya (2022), Sharma et al. (2023), Sweeney et al. (2021), Moilanen et al. (2022), Vossen et al. (2024), Ahmad et al. (2022), Xygkou et al. (2023), Ma et al. (2024), Lee and Lee (2024) |
| | Timing | When the agent should be available and how frequently it should start new interactions with the user, optionally considering preferences. | N=5 Sweeney et al. (2021), Haque and Rubya (2022), Li et al. (2024), Ahmad et al. (2022), Xygkou et al. (2023) |
| | Modality | How the interaction with the user should actually take place, optionally considering preferences. | N=8 Maharjan et al. (2022), Vossen et al. (2024), Ahmad et al. (2022), Sharma et al. (2024), Tong et al. (2023), Valentine et al. (2022), Kim et al. (2024b,a) |
| | Appearance | How the agent should appear to the user to put them at ease, optionally considering preferences. | N=3 Feijóo-García et al. (2023), Vossen et al. (2024), Lee and Lee (2024) |
| Data management | Transparency | How to let users know how their data are being handled and how proposed interventions are generated from such data. | N=3 Valentine et al. (2022), Vossen et al. (2024), Ahmad et al. (2022) |
| | Privacy | How to secure data and prevent third parties from reading. | N=5 Yang et al. (2020), Sweeney et al. (2021), Koulouri et al. (2022), Ma et al. (2024), Maharjan et al. (2022) |
| | Sensitivity Handling | How to manage sensitive and emotional data. | N=2 Capel et al. (2024), Xygkou et al. (2023) |
| Model | Selection | Which AI approach and model might be the most suitable. | N=5 Wu et al. (2024), Maharjan et al. (2022), Sharma et al. (2024), Yang et al. (2020), Sharma et al. (2023) |
| | Response Quality | Key aspects to consider to make the model give the best possible answers. | N=10 De Nieva et al. (2021), Kim et al. (2024a), Capel et al. (2024), Xygkou et al. (2023), Seo et al. (2024), Cooney et al. (2024), Li et al. (2024), Sweeney et al. (2021), Wu et al. (2024), Yang et al. (2020) |
| Study | Type | Which type of study is the most suitable to assess the validity of the proposed tool. | N=13 Moilanen et al. (2022), Kim et al. (2024a), Wu et al. (2024), Li et al. (2024), Vossen et al. (2024), Seo et al. (2024), Koulouri et al. (2022), Kim et al. (2024b), Maharjan et al. (2022), Sharma et al. (2024), Monge Roffarello and De Russis (2024), De Nieva et al. (2021), Tong et al. (2023) |
| | Participants | How many people and with what characteristics should be included in the experimental phase. | N=12 Moilanen et al. (2022), Kim et al. (2024a), Wu et al. (2024), Li et al. (2024), Vossen et al. (2024), Seo et al. (2024), Kim et al. (2024b), Maharjan et al. (2022), Sharma et al. (2024), Monge Roffarello and De Russis (2024), De Nieva et al. (2021), Tong et al. (2023) |
| | Duration | How long the experimental phase should last. | N=12 Moilanen et al. (2022), Kim et al. (2024a), Wu et al. (2024), Li et al. (2024), Vossen et al. (2024), Seo et al. (2024), Maharjan et al. (2022), Kim et al. (2024b), Sharma et al. (2024), Monge Roffarello and De Russis (2024), De Nieva et al. (2021), Tong et al. (2023) |
| | Results Validation | Which methods and control types should be applied to validate the impact of the study. | N=6 Moilanen et al. (2022), Wu et al. (2024), Li et al. (2024), Vossen et al. (2024), Sharma et al. (2024), Tong et al. (2023) |

interventions in established psychological techniques or behavioral theories, such as Cognitive-Behavioral Therapy (CBT) or Self-Determination Theory (SDT), can lend credibility and ensure scientific rigor (Haque and Rubya, 2022; De Nieva et al., 2021; Valentine et al., 2022; Kim et al., 2024a; Tong et al., 2023; Ahmad et al., 2022). A possibility is to incorporate relatable storytelling with moral insights (De Nieva et al., 2021) or encourage users to reflect deeply on their emotions and perspective (Sharma et al., 2024; De Nieva et al., 2021). Alternatively, exposing users to diverse and even contrasting suggestions could promote critical thinking and self-discovery (Kim et al., 2024a,b). Digital wellbeing tools could offer a more structured approach, like AI systems that elaborate intervention recommendations and simply present them (Valentine et al., 2022), or a more flexible one, such as LLM-based chatbots. Nevertheless, our findings show that, while users typically desire flexibility, tools like LLM-based chatbots may lead to over-dependency problems (Sharma et al., 2024). Recommendations, instead, have the capability to help the users reflect, even if the recommended interventions do not always fit with their current needs (Valentine et al., 2022). The user target and all the possible stakeholders should always be considered when defining the approach (Koulouri et al., 2022), particularly for sensitive categories like children (Seo et al., 2024). In general, it is important to design AI-based tools that demonstrate to care, and that listen and emotionally connect to the user (De Nieva et al., 2021; Xygkou et al., 2023).

#### 4.2.3. Adaptation

Adapting to the general and momentary needs of the individual user allows the delivering of interventions that are more effective in achieving the stated purpose. Developers can leverage real-time device data and suggest personalized activities, such as recommending out-of-technology experiences based on information such as accelerometer-detected inactivity (Valentine et al., 2022) or software-detected apps or websites overuse (Li et al., 2024; Tong et al., 2023), as already outlined in the *Personal Device Data* sub-dimension. However, false positives should carefully be avoided whenever possible, as this could lead to frustration (Li et al., 2024). When the AI integration includes a conversational approach, the adopted style may be dynamically adjusted from friendly to more rigorous and problem-solving-oriented when dealing with wellbeing concerns that need to be addressed (Moilanen et al., 2022). Conversational systems should be capable of recognizing and supporting users who struggle to articulate their emotions. Over-accommodation should be avoided, though, as it might interfere with professional therapeutic relationships (Kim et al., 2024a). It is important to design and implement context-sensitive intervention, with AI systems that engage only with specific topics where they can provide meaningful assistance, rather than general-purpose solutions that claim to address all issues (Young et al., 2024).

#### 4.2.4. Non-maleficence

AI-powered digital wellbeing tools should always assure appropriate and harmless outputs. Systems must not be intrusive when asking for personal data (Sharma et al., 2023), and their effectiveness should be equitable across different user demographics (Sharma et al., 2023), avoiding stereotypical assumptions, particularly for minorities and less represented user groups (Ma et al., 2024). Special care must be taken into account to prevent inappropriate recommendations for users with serious conditions (Valentine et al., 2022), and to avoid stigmatizing or criminalizing users' thoughts and behaviors (Karizat et al., 2024). Practitioners should also consider implementing measures to prevent over-reliance and even dependency on AI tools and their output (Ma et al., 2024; Seo et al., 2024), a consideration that has already emerged in the *Approach* sub-dimension, while ensuring the AI produces consistently safe, appropriate, and non-violent responses (Cooney et al., 2024; Young et al., 2024; Koulouri et al., 2022; Ma et al., 2024; Sharma et al., 2023; Seo et al., 2024; Kim et al., 2024b; Wu et al., 2024).

#### 4.2.5. Crisis management

The effectiveness of emergency handling is of major importance when designing intelligent digital wellbeing tools, as research suggests AI may be inadequate for handling severe crisis situations such as suicide or self-harm prevention (Sweeney et al., 2021). Tools such as chatbots should be programmed to identify critical behavioral patterns in conversations and share emergency numbers, but also to automatically notify mental health professionals, healthcare facilities, or loved ones (Vossen et al., 2024; Kim et al., 2024a; Haque and Rubya, 2022; Ma et al., 2024; Seo et al., 2024).

### 4.3. Interaction

The Interaction dimension provides details on Human-AI Interaction characteristics to consider during design and developing, including the AI-assumed role, tone, appearance, as well as the interaction timing and modality.

#### 4.3.1. Role

When defining the social role an AI agent can assume within wellbeing tools, the possibilities range from supportive, friend-like companions to professional-styled advisors (Ahmad et al., 2022). While some users may benefit from interacting with an AI that assumes the role of a therapeutic professional, especially when they have the possibility to select specific therapeutic approaches (Vossen et al., 2024), others might find value in more informal supportive roles, such as using the AI to simulate conversations with loved ones (Xygkou et al., 2023). Yet, these informal roles raise ethical and privacy-related questions that must be carefully addressed. As such, the design must carefully balance the potential for therapeutic benefit against the risk of emotional attachment (Capel et al., 2024), as already expressed in the *Purpose* and *Non-maleficence* sub-dimensions, while ensuring users understand that these AI interactions, regardless of the chosen role, are intended to complement, not replace, professional mental healthcare (Young et al., 2024; Cooney et al., 2024; Kim et al., 2024a; Xygkou et al., 2023), especially in serious cases where professional intervention is essential, e.g., in case of mental health support.

#### 4.3.2. Tone

Carefully defining a tone for the AI natural language expressions can improve the likability and thus the tools' potential to be supportive to users. While users generally value interacting with an AI that shows empathy and understanding (Haque and Rubya, 2022), people who seek therapy guidance in wellbeing tools tend to prefer a balanced, professional tone, specifically one that is neutral, clear, conscientious, and informative without being verbose (Moilanen et al., 2022). To find a trade-off, the design can offer users the flexibility to select their preferred interaction tone or have AI agents that automatically adapt to what users seem to need (Vossen et al., 2024; Ahmad et al., 2022), in align with the *Adaptation* sub-dimension. In doing so, however, practitioners should consider that highly emotive or human-like tones may inadvertently raise privacy concerns, as users might feel more judged when the AI appears to possess thoughts and emotions (Lee and Lee, 2024). The non-judgmental nature of AI interactions represents indeed a unique advantage to value (Xygkou et al., 2023), and situations where this tone is combined with improvements in the AI's ability to mimic genuine emotional expression can enhance user trust and engagement (Sweeney et al., 2021). This suggests the design opportunity to create adaptive interaction tones that create a safe, comfortable environment for users (Ma et al., 2024).

#### 4.3.3. Timing

Temporal dynamics patterns range from constant availability to strategically timed interventions. While users often express a preference for continuous availability, appreciating the “tireless” nature of AI compared to human support (Xygkou et al., 2023; Haque and Rubya, 2022), the tools can also be designed to intervene at specific moments (Li et al., 2024). Indeed, proactive interaction, including scheduled check-ins and reminders, has received positive evaluation from mental health experts (Sweeney et al., 2021), suggesting that practitioners can create systems that combine always-on accessibility with customizable proactive outreach based on user preferences (Ahmad et al., 2022).

#### 4.3.4. Modality

In designing and implementing AI-powered digital wellbeing tools, multiple interaction modalities can be considered, accommodating different preferences and situational needs. When opting for conversational assistants that leverage a voice-based interaction modality, it is important to consider that the appropriateness of voice interaction can vary significantly depending on environmental factors. For example, some users might struggle to find private spaces to talk, while others might benefit from voice interactions being overheard to potentially include others nearby in the conversation (Maharjan et al., 2022). Practitioners can also implement systems that allow users to seamlessly switch between voice and text-based interactions based on their current preferences and situations (Ahmad et al., 2022). Text-based interactions can be further customized by allowing users to adjust parameters like typing speed to match their comfort level (Vossen et al., 2024). Conversational agents are not the only interaction modality to explore, despite being the most diffused. Practitioners can also consider building systems that guide users through a more structured approach, such as tools that give users a series of choices to help them detail their current wellbeing struggle and offer personalized supportive paths (Sharma et al., 2024), recommender systems that leverage real-time data to propose a set of interventions adapted to the user needs (Wu et al., 2024; Tong et al., 2023; Valentine et al., 2022), or writing assistants to support introspective reflection and emotional expression (Kim et al., 2024b,a).

#### 4.3.5. Appearance

Appearance describes choices to take when defining the aesthetic look the AI agent shows to the user. Practitioners can implement customizable avatar systems (Vossen et al., 2024), with particular attention to demographic characteristics, as users tend to prefer agents that share their age, gender, ethnicity, accent, and cultural background (Feijóo-García et al., 2023). Users generally respond more positively to anthropomorphic profile images than to robotic ones (Lee and Lee, 2024). Therefore, there are opportunities to create either highly personalized agent appearances that mirror user characteristics or offer a range of customization options that allow users to select their preferred agent representation.

### 4.4. Data management

The Data Management dimension discusses the sensitive topic of how to collect and manage user information, particularly intimate ones, while assuring transparency over such collection and management.

#### 4.4.1. Transparency

Implementing transparent data management in digital wellbeing applications that exploit AI can range from revealing the reasoning behind AI-generated recommendations and their data sources (Valentine et al., 2022; Ahmad et al., 2022) to empowering users with granular control over their information handling preferences (Vossen et al., 2024). Automated, yet transparent, data collection can streamline the user experience and, in some cases, increase the quality and personalization of the outputs. Practitioners may consider to implement features that give user control over their data, like selective data history, to allow them to influence future recommendations (Valentine et al., 2022).

#### 4.4.2. Privacy

Encrypted conversations stored on remote servers (Yang et al., 2020) can preserve privacy. Similarly, privacy-protecting measures may involve the definition of clear protocols to define who can access what information, even among researchers (Maharjan et al., 2022). Given the importance of protecting user data (Koulouri et al., 2022; Sweeney et al., 2021), measures can be implemented to prevent continuous listening by voice agents, ensure explicit consent mechanisms for data sharing, and establish robust security measures to prevent unauthorized third-party access to personal information (Ma et al., 2024).

#### 4.4.3. Sensitivity handling

Practitioners should address with particular caution intimate and emotional data (Capel et al., 2024), especially when this data is handled by wellbeing apps that deal with mental health. In this regard, it is not only important to carefully define who can access that data, but also to raise fundamental ethic questions, such as whether it is appropriate for users to discuss sensitive personal information about others with AI agents, even if they need to do it for their digital wellbeing therapeutic path. For example in Xygkou et al. (2023), study participants found it therapeutic talking with a chatbot that they instructed to pretend being a deceised loved one to cope with grief, but to make this approach effective they needed to share with the agents personal information about their dear one, potentially against their consent.

### 4.5. Model

The Model dimension offers possibilities to reflect on selecting the right AI model when implementing digital wellbeing tools, in order to ensure high response quality and avoid frustration and tool abandonment.

#### 4.5.1. Selection

The selection of a proper AI model represents a critical decision when designing for digital wellbeing, and it presents diverse architectural choices, each with distinct advantages. When working with models that respond using natural language, specialized conversational frameworks can provide structured interaction patterns under the control of developers (Maharjan et al., 2022), while LLMs, instead, have sophisticated dialogue capabilities, but are also hallucination-prone (Sharma et al., 2024, 2023). Our framework also highlights the possibility to leverage hybrid approaches like the one proposed by Yang et al. (2020) that developed a chatbot supporting college students mental health. In particular, they used input keywords to know when to switch between questioning a deep learning based language model to support emotional concerns with flexibility, and a database of predefined outputs that more rigorously and effectively address difficulties related to academic study. Regardless of the chosen AI type, it is important to consider using open-source models, as they offer enhanced data privacy protections (Wu et al., 2024). Overall, the selection of an AI model should consider aspects such as privacy, conversational depth, domain expertise, and implementation complexity.

#### 4.5.2. Response quality

To ensure effective response quality, misunderstandings, inaccurate responses, lack of coherence between outputs and repetitive interactions should be prevented, as they are sources of user frustration and tool abandonment (De Nieva et al., 2021; Kim et al., 2024a; Xygkou et al., 2023; Seo et al., 2024; Li et al., 2024). Clear indicators highlighting when the system’s confidence is low are beneficial in that sense (Wu et al., 2024). When developing AI-based systems for a certain target, it is fundamental to ensure the model has some knowledge of the specific cultural context of the target group, and to carefully alert users of the model limitation if this prior knowledge is missing, e.g., when using third-party general purpose LLMs (Capel et al., 2024). Finally, providing users with meaningful interactions with AI also involves demonstrating the ability to emulate empathy (Sweeney et al., 2021) and identify information that is truly important for the user's specific wellbeing path (Cooney et al., 2024).

### 4.6. Study

This dimension provides insight into how an empirical study regarding AI tools for digital wellbeing should be conducted, considering the implementation context and the aim of the research, to obtain more valid results.

#### 4.6.1. Type

Based on the examined studies, practitioners have several methodological approaches available for evaluating AI-powered digital wellbeing tools. They can choose between laboratory-controlled environments (Moilanen et al., 2022; Seo et al., 2024; Koulouri et al., 2022) and "in the wild" field experiments (De Nieva et al., 2021; Monge Roffarello and De Russis, 2024; Sharma et al., 2024; Maharjan et al., 2022; Kim et al., 2024b; Vossen et al., 2024; Li et al., 2024; Wu et al., 2024; Kim et al., 2024a), each offering distinct insights. Field studies can involve extended usage periods and incorporate multiple data collection methods such as usage tracking or periodic questionnaires. While for laboratory studies, practitioners may operate means such as semi-structured interviews or assisted sessions. Other means like psychological scale assessments, app usage frequency, and session duration tracking, may be valid for both types of study. Some practitioners can also prefer qualitative feedback and therapeutic relationship assessment. The diversity of these methodological approaches suggests that practitioners can flexibly design studies to capture nuanced user experiences and tool effectiveness across different contexts and digital wellbeing scenarios.

#### 4.6.2. Participants

When designing AI-powered digital wellbeing tools, researchers should consider a wide range of potential participant demographics. The experimental phase can consider diverse population segments, focusing on various factors, such as age, profession or habits. Studies have demonstrated the benefits of involving varied age groups, ranging from youth (Seo et al., 2024) to young adults (Maharjan et al., 2022), as well as different occupations, such as high school (De Nieva et al., 2021) and university students (Kim et al., 2024b), and workers (Tong et al., 2023). In terms of participant recruitment, participant pools that originate from specific regions (Seo et al., 2024) are convenient for methodologies that necessitate in-person interaction between researchers and testers. Nevertheless, it is important to highlight that such an approach reduces the generalizability of the results, being a factor to consider when choosing the experimental population with respect to the target population of the tool. The generalizability of the results is also influenced by the number of selected subjects. Such a number can vary from very few participants (Monge Roffarello and De Russis, 2024) to larger populations, from dozens of participants (Kim et al., 2024a; Wu et al., 2024; Li et al., 2024; Vossen et al., 2024; Seo et al., 2024; Kim et al., 2024b; Maharjan et al., 2022; De Nieva et al., 2021; Tong et al., 2023) to hundreds (Moilanen et al., 2022) or even thousands (Sharma et al., 2023), depending on the available resources. While studies involving a small population may allow for more rigor, the few participants might compromise the generalizability of the results. Conversely, studies involving bigger populations tend to be less detailed, providing less specific but more generalizable results. This means that considering the resources available and the chosen study type, practitioners, to assess the validity and generalizability of their study, should reflect on how to ensure the proper number of participants that allow them to obtain the level of detail needed for the results.

#### 4.6.3. Duration

When evaluating AI-powered digital wellbeing tools, researchers have explored a diverse range of experimental durations that reflect the complexity of habit formation and technology-mediated psychological interventions. The literature reveals experimental periods that range from brief interactions of approximately 30–40 minutes (Moilanen et al., 2022; Seo et al., 2024) to more complex studies spanning 1–2 weeks (Vossen et al., 2024; Kim et al., 2024b; De Nieva et al., 2021). Moreover, other researchers have pursued more extended investigations ranging from converging around 3–5 week periods (Kim et al., 2024a; Wu et al., 2024; Li et al., 2024; Maharjan et al., 2022; Tong et al., 2023; Monge Roffarello and De Russis, 2024), to comprehensive 9-month longitudinal designs (Sharma et al., 2023). While the study duration primarily depends on the evaluated tool, we must highlight that longer periods might be necessary to effectively assess habit acquisition and intervention efficacy, particularly when exploring nuanced psychological support mechanisms (Monge Roffarello and De Russis, 2024).

#### 4.6.4. Results validation

As practitioners in the field of AI for digital wellbeing may choose to explore numerous experimental design methodologies, the results obtained consequently should be validated through some precise strategy for assuring their reliability and validity. It is possible to experiment with design variations among different groups of subjects or within the same, ranging from controlling the presence or absence of specific technological features like large language models (LLMs) and reflective prompts (Wu et al., 2024; Li et al., 2024), to manipulating interaction modalities such as randomized versus fully customizable chatbot configurations (Vossen et al., 2024). Practitioners can carry out comparative studies that employ control groups with different characteristics, including subjects with minimal interventions, like weekly stress level surveys (Tong et al., 2023) or varied informational content availability (Sharma et al., 2024). Depending on the study and on the available amount of participants, a within-subject approach (as seen in Section 3), making all users try all variations and analyzing their preferences and responses, might be more appropriate. It is interesting that the choice of a validation method may be related to the specific focus of wellbeing considered. For example, while generally it seems to be more frequent to have between-subjects validation, from Monge Roffarello and De Russis (2023) it emerges that, for technology overuse related wellbeing problems, many researchers use a within-subjects approach. So, practitioners should carefully consider which type of validation they should apply to their study and try to choose the best option to assess the validity of their results.

## 5. Discussion

Our systematic literature review revealed interesting challenges and open problems in the field of AI-powered digital wellbeing applications. By analyzing and synthesizing the papers in our *General Corpus* (Section 3), we developed a framework comprising 6 key dimensions and 23 sub-dimensions (Section 4) that represents a reflective tool for researchers and developers who are leveraging AI to support people's digital wellbeing.

The primary objective of our framework is not to prescribe a rigid set of rules but rather to provide a flexible guide that encourages a thoughtful approach to developing AI-integrated digital wellbeing applications. While the framework provides structured guidance, practitioners should consider the practical constraints of each case, such as available resources or technological limitations. As such, we do not mean the framework as a protocol to be enforced but rather as a flexible tool that can be interpreted and applied according to specific research or development scenarios. The framework functions as an analytical lens, helping researchers and developers explore and address the multifaceted aspects of digital wellbeing interventions.

Central to our framework is a user-centric approach. Recognizing the sensitive nature of wellbeing interventions, we emphasize the critical importance of placing the user at the core of any design process. Namely, it is important to deeply understand and respect the target population's specific characteristics, needs, perceptions, and contexts. Each application should be meticulously tailored to its intended users, acknowledging that digital wellbeing is not a one-size-fits-all concept but a profoundly personal and contextual experience.

The framework offers three main advantages. First, it serves as a checklist, ensuring researchers recognize critical dimensions. Second, it provides a structured approach to exploring potential challenges and opportunities in AI-powered digital wellbeing applications. Third, the framework can inspire innovative approaches and reveal unexplored research directions.

An essential aspect of our framework is its potential to promote more responsible and ethical development of digital wellbeing technologies. Encouraging reflection on aspects like user privacy, potential risks, and the psychological impact of interventions helps mitigate potential unintended negative consequences of AI-powered applications.

However, as we will discuss in Section 5.2, it is crucial to acknowledge that our framework is not without limitations. The rapid evolution of AI technologies means that researchers using this framework should view it as a dynamic tool that requires reassessment and adaptation to contexts and emerging technologies and problems.

### *5.1. Using the framework in practice: a case study*

To demonstrate the utility of our proposed design framework and clarify its application in real-world scenarios, we present a case study involving the development of an AI-powered digital wellbeing application. This case study translates the framework into practice, presenting design decisions through a high-fidelity prototype developed in Figma (Figma, 2025).

The prototype implements an intelligent intervention tool designed to help students reduce excessive technology usage during study-time. By leveraging LLMs, the application provides personalized and adaptive guidance to help students stay focused while studying and gradually develop autonomy from the application itself. The translation of this objective into design choices is systematically guided by our framework's dimensions, with decisions visually represented in the prototype interfaces of Fig. 6. Table 7 provides a comprehensive overview of how each framework dimension has been implemented in the prototype, offering an overview of the relationship between framework and case study design.

The prototype demonstrates how the *User Information* dimension is implemented through a carefully designed onboarding interface that minimizes invasive data collection while gathering essential information. The user input forms collect basic *Demographics* such as name and study area, as shown in Fig. 6(a), alongside *Lifestyle* information including hobbies and free time activities. The prototype collects user-defined *Goals* for reducing device distractions and implements transparent collection of *Personal Device Metrics* with minimal cognitive overloading.

The *Intervention* dimension is translated into the prototype's main functionalities, which have the *Purpose* of promoting behavioral change towards healthier technology habits aligned with user-defined goals. As the main *Approach*, the interface implements daily micro-tasks designed to progressively increase user autonomy. Fig. 6(b) shows the home page of the prototype where the daily micro-tasks can be seen alongside with user progress bar. The prototype demonstrates *Adaptation* of the intervention supported by an LLM providing progressively challenging tasks at first and later decreasing rigidity as user shows improvement. The design ensures *Non-Maleficence* through a non-judgmental and encouraging interface that account for user feelings throughout the intervention. No specific emergency is foreseen and therefore no *Crisis Management* is implemented.

The *Interaction* dimension is embodied in the prototype's interface design, where the app assumes a supportive *Role* through its cheering messages, as the one in Fig. 6(a). While the main intervention *Modality* of the app consists of suggestion-based daily micro-interventions, these are complemented by optional chat support for emotional guidance. As for the *Appearance* of the chat-interlocutor, the prototype features a cartoon-style avatar that maintains age-appropriate relatability without hyper-realistic representation. Concerning *Timing*, the interface design ensures chat consistent availability while providing daily progress summaries and customizable *Tone* based on user preferences.

As for the *Data Management* dimension, the prototype interface design focuses on *Transparency* for data handling, with clear communication about data collection and usage purposes. The future application will implement minimal data sharing through third-party APIs to ensure users *Privacy*, with sensitive data stored locally to assure a proper *Sensitivity Handling*. The prototype's design ensures users understand how their information are used to supports their personal goals while maintaining data security.

Concerning the last two dimensions, we have not actually implemented them since we only developed a prototype, anyway we now present a possible future implementation of the two. For the *Model* dimension we foresee the *Selection* of GPT-3.5 or GPT-4 via external APIs, with prompts enriched by evidence-based micro-intervention strategies from established literature (Lyngs et al., 2019; Roffarello and De Russis, 2023) to improve *Response Quality*. This approach ensures the application's responses maintain scientific validity while providing personalized guidance.

Finally, a possible *Study* for the evaluation of the proposed case study, a corresponding implemented application can be tested through an "in the wild" assessment, for the *Type* subdimension, with a *Duration* of at least 2 to 3 months to investigate long-term change of habits. A minimum of 20 *Participants* should be selected among different student populations. They should try personalized versus randomized interventions to provide a strong *Result Validation* where the effects of LLM-integration are clear and distinguishable.

### *5.2. Limitations*

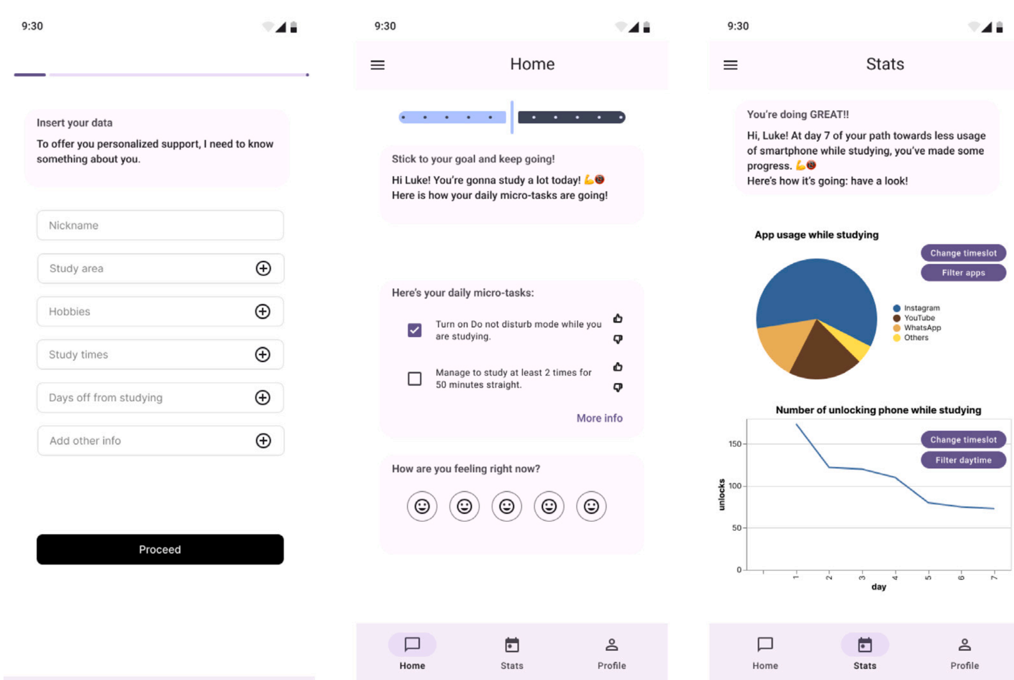
Our work and the proposed framework have limitations that could inform future research in this critical and evolving field.

In our systematic literature review, the applied time-based filter potentially excludes valuable emerging publications. As the use of AI for digital wellbeing represents a rapidly evolving research domain, new papers published after our review might contain critical insights that our current methodology has not captured. The keyword selection process, that was focused on explicit terms like "digital wellbeing" and "mental health", may have inadvertently overlooked relevant research with more generic terminology. This approach narrowed our research scope, potentially missing valuable contributions that address digital wellbeing concepts less directly. Furthermore, the perspective of our review was predominantly anchored in Computer Science, Computer Engineering, and Human-Computer Interaction. Consequently, potentially valuable interdisciplinary perspectives, particularly from fields like psychology, sociology, and health sciences, may not have been represented. The main purpose of this work is to provide guidance to those working in the field of HCI. It can be used alongside, but cannot replace, other works that focus on more specific aspects of other research fields. This frames a significant avenue for future research, encouraging cross-disciplinary collaboration and broader knowledge integration.

Regarding the generalizability of our framework, while the reviewed studies demonstrated diversity in participant demographics, such as variations in age, nationality, and individual conditions, we acknowledge the potential for underrepresentation or exclusion of specific user groups. We therefore encourage future researchers to critically examine and expand upon our framework, particularly by conducting targeted studies that enhance the representation of currently marginalized or overlooked populations.

**Table 7**
Application of the design framework to the case study.

| Dimension | Sub-dimension | Case study application |
|---|---|---|
| User information | Demographics | Basic data such as name and study area are collected minimally. |
| | Lifestyle | Includes hobbies and free time activities outside studying. |
| | Goals | User-defined goals to reduce personal device overuse while studying. |
| | Personal Device Metrics | Automatically collected smartphone usage data, with transparency and low onboarding effort. |
| Intervention | Purpose | Promote healthier habits aligned with user-defined goals. |
| | Approach | Daily micro-interventions to gradually build user autonomy. |
| | Adaptation | Adaptive interventions over time, supported by AI. |
| | Non-Maleficence | Ensures a non-judgmental, safe user experience. |
| | Crisis Management | No crisis management planned, but user well-being is prioritized. |
| Interaction | Role | Supportive digital companion. |
| | Tone | Customizable emotional tone based on user preferences. |
| | Timing | Always available with daily progress summaries. |
| | Modality | Suggestion-based interface with optional emotional chat support. |
| | Appearance | Cartoon-style avatar, age-relevant but not hyper-realistic. |
| Data management | Transparency | Clear communication on data usage and purpose. |
| | Privacy | Use of third-party APIs with minimal data sharing. |
| | Sensitivity Handling | Only necessary sensitive data collected, stored locally. |
| Model | Selection | Integration with GPT-3.5 or GPT-4 via external APIs. |
| | Response Quality | Prompts enriched with evidence from existing literature. |
| Study | Type | Long-term "in the wild" study. |
| | Participants | At least 20 students with diverse profiles. |
| | Duration | Several months to assess behavioral change. |
| | Results Validation | Comparison between personalized and randomized interventions. |




(a) User information form. (b) Home page. (c) Statistics.

**Fig. 6.** Example of pages from the Figma prototype of the case study.

## 6. Conclusion

This study has addressed a critical gap in developing AI-powered digital wellbeing applications by proposing a structured design framework. Our systematic literature review and subsequent analysis revealed the complexities of digital wellbeing tools, highlighting potential and significant challenges in leveraging artificial intelligence to support human psychological and emotional wellness.

The proposed framework, encompassing 6 critical dimensions and 23 subdimensions, provides researchers and developers with a robust, ethically grounded approach to creating digital tools that do respect users and their interests. By emphasizing key considerations such as user data privacy, intervention strategies, and personalization, we aim to mitigate potential risks associated with poorly designed AI applications, such as over-reliance or biased responses, while maximizing their positive impact.

Future research and efforts should focus on implementing and testing our framework across diverse digital wellbeing applications. Refining and expanding this framework would also be possible considering diverse research perspectives. Fostering collaboration between interdisciplinary stakeholders, the field can continue to advance towards the ethical and effective use of AI to enhance digital wellbeing. Furthermore, over time it will be important to address the future advancements in AI that will pose new challenges.

As digital technologies integrate more deeply into human experience, the need for thoughtful, user-centric design becomes increasingly paramount. This framework represents a step towards more responsible and effective digital wellbeing solutions, balancing technological innovation with fundamental human needs and ethical considerations.

## CRediT authorship contribution statement

**Luca Scibetta:** Writing – review & editing, Writing – original draft, Visualization, Project administration, Methodology, Investigation, Formal analysis, Conceptualization. **Massimiliano Pellegrino:** Writing – review & editing, Writing – original draft, Visualization, Project administration, Methodology, Investigation, Formal analysis, Conceptualization. **Alberto Monge Roffarello:** Writing – review & editing, Visualization, Supervision, Project administration, Methodology, Investigation, Formal analysis, Conceptualization. **Luigi De Russis:** Writing – review & editing, Visualization, Supervision, Project administration, Methodology, Investigation, Formal analysis, Conceptualization.

## Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT and ClaudeAI in order to improve text quality and fluency. After using these tools, the authors reviewed end edited the content as needed and take full responsibility for the content of the publication.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

No data was used for the research described in the article.

## References

Abd-Alrazaq, A.A., Alajlani, M., Ali, N., Denecke, K., Bewick, B.M., Househ, M., 2021. Perceptions and opinions of patients about mental health chatbots: scoping review. J. Med. Internet Res. 23 (1), e17828.

Ahmad, R., Siemon, D., Gnewuch, U., Robra-Bissantz, S., 2022. Designing personality-adaptive conversational agents for mental health care. Inf. Syst. Front. 24 (3), 923–943. http://dx.doi.org/10.1007/s10796-022-10254-9.

Bandi, A., Adapa, P.V.S.R., Kuchi, Y.E.V.P.K., 2023. The power of generative ai: A review of requirements, models, input–output formats, evaluation metrics, and challenges. Futur. Internet 15 (8), 260.

Bavaresco, R., Ren, Y., Barbosa, J., Li, G., 2024. An ontology-based framework for worker's health reasoning enabled by machine learning. Comput. Ind. Eng. 193 (C), http://dx.doi.org/10.1016/j.cie.2024.110310.

Capel, T., Ploderer, B., Bircanin, F., Hanmer, S., Yates, J.P., Wang, J., Khor, K.L., Leong, T.W., Wadley, G., Newcomb, M., 2024. Studying self-care with generative AI tools: Lessons for design. In: Proceedings of the 2024 ACM Designing Interactive Systems Conference. DIS '24, Association for Computing Machinery, New York, NY, USA, pp. 1620–1637. http://dx.doi.org/10.1145/3643834.3661614.

Casu, M., Triscari, S., Battiato, S., Guarnera, L., Caponnetto, P., 2024. AI chatbots for mental health: A scoping review of effectiveness, feasibility, and applications. Appl. Sci. 14 (13), 5889.

Chen, T., Shen, Y., Chen, X., Zhang, L., 2024. PsyChatbot: A psychological counseling agent towards depressed Chinese population based on cognitive behavioural therapy. ACM Trans. Asian Low-Resour. Lang. Inf. Process. http://dx.doi.org/10.1145/3676962, Just Accepted.

Cooney, O., Doherty, K., Barry, M., Coyle, D., Doherty, G., Cowan, B.R., 2024. Using speech agents for mood logging within blended mental healthcare: Mental healthcare practitioners' perspectives. In: Proceedings of the 6th ACM Conference on Conversational User Interfaces. CUI '24, Association for Computing Machinery, New York, NY, USA, http://dx.doi.org/10.1145/3640794.3665540.

De Nieva, J.O., Joaquin, J.A., Tan, C.B., Marc Te, R.K., Ong, E., 2021. Investigating students' use of a mental health chatbot to alleviate academic stress. In: 6th International ACM in-Cooperation HCI and UX Conference. CHIuXiD '20, Association for Computing Machinery, New York, NY, USA, pp. 1–10. http://dx.doi.org/10.1145/3431656.3431657.

Elnawawy, M., Hallajiyan, M., Mitra, G., Iqbal, S., Pattabiraman, K., 2024. Systematically assessing the security risks of AI/ML-enabled connected healthcare systems. In: 2024 IEEE/ACM Conference on Connected Health: Applications, Systems and Engineering Technologies. CHASE, IEEE Press, pp. 97–108. http://dx.doi.org/10.1109/CHASE60773.2024.00019.

Feijóo-García, P.G., Wrenn, C., Gomes de Siqueira, A., Ghosh, R., Stuart, J., Yao, H., Lok, B., 2024. Exploring the effects of user-agent and user-designer similarity in virtual human design to promote mental health intentions for college students. ACM Trans. Appl. Percept. http://dx.doi.org/10.1145/3689822, Just Accepted.

Feijóo-García, P.G., Wrenn, C., Stuart, J., De Siqueira, A.G., Lok, B., 2023. Participatory design of virtual humans for mental health support among North American computer science students: Voice, appearance, and the similarity-attraction effect. ACM Trans. Appl. Percept. 20 (3), http://dx.doi.org/10.1145/3613961.

Figma, 2025. About page. URL https://www.figma.com/about/, (Accessed 03 June 2025).

Floridi, L., 2014. The Fourth Revolution. Oxford University Press, Oxford.

Gama, F., Tyskbo, D., Nygren, J., Barlow, J., Reed, J., Svedberg, P., 2022. Implementation frameworks for artificial intelligence translation into health care practice: Scoping review. J Med Internet Res 24 (1), e32215. http://dx.doi.org/10.2196/32215, URL https://www.jmir.org/2022/1/e32215.

Gobin-Rahimbux, B., Gooda Sahib, N., Peerthy, N., Taylor, A., 2023. A voice-based personal assistant for mental health in kreol morisien. J. Electr. Comput. Eng. 2023 (1), 5532967. http://dx.doi.org/10.1155/2023/5532967, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1155/2023/5532967.

Gross, N., 2023. What ChatGPT tells us about gender: A cautionary tale about performativity and gender biases in AI. Soc. Sci. URL https://api.semanticscholar.org/CorpusID:260600031.

Haque, M.R., Rubya, S., 2022. "For an app supposed to make its users feel better, it sure is a joke" - an analysis of user reviews of mobile mental health applications. Proc. ACM Hum.-Comput. Interact. 6 (CSCW2), http://dx.doi.org/10.1145/3555146.

Hoare, E., Milton, K., Foster, C., Allender, S., 2016. The associations between sedentary behaviour and mental health among adolescents: a systematic review. Int. J. Behav. Nutr. Physical Act. 13, http://dx.doi.org/10.1186/s12966-016-0432-4.

Huang, Y., Li, L., Gan, Y., Wang, C., Jiang, H., Cao, S., Lu, Z., 2020. Sedentary behaviors and risk of depression: a meta-analysis of prospective studies. Transl. Psychiatry 10, http://dx.doi.org/10.1038/s41398-020-0715-z.

Jingar, M., Lindgren, H., 2019. Tangible communication of emotions with a digital companion for managing stress: An exploratory co-design study. In: Proceedings of the 7th International Conference on Human-Agent Interaction. HAI '19, Association for Computing Machinery, New York, NY, USA, pp. 28–36. http://dx.doi.org/10.1145/3349537.3351907.

Karizat, N., Vinson, A.H., Parthasarathy, S., Andalibi, N., 2024. Patent applications as glimpses into the sociotechnical imaginary: Ethical speculation on the imagined futures of emotion AI for mental health monitoring and detection. Proc. ACM Hum.-Comput. Interact. 8 (CSCW1), http://dx.doi.org/10.1145/3637383.

Kim, T., Bae, S., Kim, H.A., Lee, S.-W., Hong, H., Yang, C., Kim, Y.-H., 2024a. MindfulDiary: Harnessing large language model to support psychiatric patients' journaling. In: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems. CHI '24, Association for Computing Machinery, New York, NY, USA, http://dx.doi.org/10.1145/3613904.3642937.

Kim, T., Shin, D., Kim, Y.-H., Hong, H., 2024b. DiaryMate: Understanding user perceptions and experience in human-ai collaboration for personal journaling. In: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems. CHI '24, Association for Computing Machinery, New York, NY, USA, http://dx.doi.org/10.1145/3613904.3642693.

Kip, H., Beerlage-de Jong, N., van Gemert-Pijnen, L.J., Kelders, S.M., 2025. The CeHRes roadmap 2.0: update of a holistic framework for development, implementation, and evaluation of ehealth technologies. J. Med. Internet Res. 27, e59601.

Koulouri, T., Macredie, R.D., Olakitan, D., 2022. Chatbots to support Young adults' mental health: An exploratory study of acceptability. ACM Trans. Interact. Intell. Syst. 12 (2), http://dx.doi.org/10.1145/3485874.

Lee, J., Lee, D., 2024. User perception and self-disclosure towards an AI psychotherapy chatbot according to the anthropomorphism of its profile picture. Telemat. Inf. 85 (C), http://dx.doi.org/10.1016/j.tele.2023.102052.

Lee, M.H., Montgomery, J.M., Lai, C.K., 2024. Large language models portray socially subordinate groups as more homogeneous, consistent with a bias observed in humans. In: Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency. URL https://api.semanticscholar.org/CorpusID:267028270.

Lekadir, K., Frangi, A.F., Porras, A.R., Glocker, B., Cintas, C., Langlotz, C.P., Weicken, E., Asselbergs, F.W., Prior, F., Collins, G.S., Kaissis, G., Tsakou, G., Buvat, I., Kalpathy-Cramer, J., Mongan, J., Schnabel, J.A., Kushibar, K., Riklund, K., Marias, K., Amugongo, L.M., Fromont, L.A., Maier-Hein, L., Cerdá-Alberich, L., Martí-Bonmatí, L., Cardoso, M.J., Bobowicz, M., Shabani, M., Tsiknakis, M., Zuluaga, M.A., Fritzsche, M.-C., Camacho, M., Linguraru, M.G., Wenzel, M., De Bruijne, M., Tolsgaard, M.G., Goisauf, M., Cano Abadía, M., Papanikolaou, N., Lazrak, N., Pujol, O., Osuala, R., Napel, S., Colantonio, S., Joshi, S., Klein, S., Aussó, S., Rogers, W.A., Salahuddin, Z., Starmans, M.P.A., 2025. FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare. Br. Med. J. 388, http://dx.doi.org/10.1136/bmj-2024-081554, arXiv:https://www.bmj.com/content/388/bmj-2024-081554.full.pdf, URL https://www.bmj.com/content/388/bmj-2024-081554.

Li, Z., Liang, M., Lc, R., Luo, Y., 2024. StayFocused: Examining the effects of reflective prompts and chatbot support on compulsive smartphone use. In: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems. CHI '24, Association for Computing Machinery, New York, NY, USA, http://dx.doi.org/10.1145/3613904.3642479.

Lyngs, U., Lukoff, K., Slovak, P., Binns, R., Slack, A., Inzlicht, M., Van Kleek, M., Shadbolt, N., 2019. Self-control in cyberspace: Applying dual systems theory to a review of digital self-control tools. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. pp. 1–18.

Lyngs, U., Lukoff, K., Slovak, P., Inzlicht, M., Freed, M., Andrews, H., Tinsman, C., Csuka, L., Alberts, L., Oldemburgo De Mello, V., Makransky, G., Hornbæk, K., Van Kleek, M., Shadbolt, N., 2024. "I finally felt I had the tools to control these urges": Empowering students to achieve their device use goals with the reduce digital distraction workshop. In: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems. CHI '24, Association for Computing Machinery, New York, NY, USA, http://dx.doi.org/10.1145/3613904.3642946.

Ma, Z., Mei, Y., Long, Y., Su, Z., Gajos, K.Z., 2024. Evaluating the experience of lgbtq+ people using large language model based chatbots for mental health support. In: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems. CHI '24, Association for Computing Machinery, New York, NY, USA, http://dx.doi.org/10.1145/3613904.3642482.

Maharjan, R., Doherty, K., Rohani, D.A., Bækgaard, P., Bardram, J.E., 2022. Experiences of a speech-enabled conversational agent for the self-report of well-being among people living with affective disorders: An in-the-wild study. ACM Trans. Interact. Intell. Syst. 12 (2), http://dx.doi.org/10.1145/3484508.

Michie, S., Van Stralen, M.M., West, R., 2011. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement. Sci. 6 (1), 42.

Moilanen, J., Visuri, A., Suryanarayana, S.A., Alorwu, A., Yatani, K., Hosio, S., 2022. Measuring the effect of mental health chatbot personality on user engagement. In: Proceedings of the 21st International Conference on Mobile and Ubiquitous Multimedia. MUM '22, Association for Computing Machinery, New York, NY, USA, pp. 138–150. http://dx.doi.org/10.1145/3568444.3568464.

Monge Roffarello, A., De Russis, L., 2019. The race towards digital wellbeing: Issues and opportunities. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. CHI '19, Association for Computing Machinery, New York, NY, USA, pp. 1–14. http://dx.doi.org/10.1145/3290605.3300616.

Monge Roffarello, A., De Russis, L., 2024. Hey StepByStep! can you teach me how to use my phone better? Int. J. Hum.-Comput. Stud. 183 (C), http://dx.doi.org/10.1016/j.ijhcs.2023.103195.

Mummah, S.A., Robinson, T.N., King, A.C., Gardner, C.D., Sutton, S., 2016. IDEAS (integrate, design, assess, and share): a framework and toolkit of strategies for the development of more effective digital interventions to change health behavior. J. Med. Internet Res. 18 (12), e317.

Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., Chou, R., Glanville, J., Grimshaw, J.M., Hróbjartsson, A., Lalu, M.M., Li, T., Loder, E.W., Mayo-Wilson, E., McDonald, S., McGuinness, L.A., Stewart, L.A., Thomas, J., Tricco, A.C., Welch, V.A., Whiting, P., Moher, D., 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Syst. Rev. 10 (1), 89. http://dx.doi.org/10.1186/s13643-021-01626-4.

Parry, D.A., le Roux, D.B., Morton, J., Pons, R., Pretorius, R., Schoeman, A., 2023. Digital wellbeing applications: Adoption, use and perceived effects. Comput. Hum. Behav. 139, 107542. http://dx.doi.org/10.1016/j.chb.2022.107542, URL https://www.sciencedirect.com/science/article/pii/S0747563222003624.

Pellegrino, G., Gervasi, M., Angelelli, M., Corallo, A., 2024. A conceptual framework for digital twin in healthcare: Evidence from a systematic meta-review. Inf. Syst. Front. 1–26.

Rajashekar, N.C., Shin, Y.E., Pu, Y., Chung, S., You, K., Giuffre, M., Chan, C.E., Saarinen, T., Hsiao, A., Sekhon, J., Wong, A.H., Evans, L.V., Kizilcec, R.F., Laine, L., Mccall, T., Shung, D., 2024. Human-algorithmic interaction using a large language model-augmented artificial intelligence clinical decision support system. In: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems. CHI '24, Association for Computing Machinery, New York, NY, USA, http://dx.doi.org/10.1145/3613904.3642024.

Roffarello, A.M., De Russis, L., 2023. Achieving digital wellbeing through digital self-control tools: A systematic review and meta-analysis. ACM Trans. Comput.-Hum. Interact. 30 (4), http://dx.doi.org/10.1145/3571810.

Seo, W., Yang, C., Kim, Y.-H., 2024. ChaCha: Leveraging large language models to prompt children to share their emotions about personal events. In: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems. CHI '24, Association for Computing Machinery, New York, NY, USA, http://dx.doi.org/10.1145/3613904.3642152.

Sharma, R., Mirzakhalov, J., Bharti, P., Goyal, R., Schmidt, T., Chellappan, S., 2023. A friend in need is a friend indeed: Investigating the quality of training data from peers for auto-generating empathetic textual responses to non-sensitive posts in a cohort of college students. ACM J. Comput. Sustain. Soc. 1 (2), http://dx.doi.org/10.1145/3616382.

Sharma, A., Rushton, K., Lin, I.W., Nguyen, T., Althoff, T., 2024. Facilitating self-guided mental health interventions through human-language model interaction: A case study of cognitive restructuring. In: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems. CHI '24, Association for Computing Machinery, New York, NY, USA, http://dx.doi.org/10.1145/3613904.3642761.

Smits, M., Kim, C.M., van Goor, H., Ludden, G.D.S., 2022. From digital health to digital well-being: Systematic scoping review. J Med Internet Res 24 (4), e33787. http://dx.doi.org/10.2196/33787, URL https://www.jmir.org/2022/4/e33787.

Stefanidi, E., 2022. Improving the digital well-being of neurotypical & neurodivergent children considering their care ecosystem. In: Proceedings of the 21st Annual ACM Interaction Design and Children Conference. IDC '22, Association for Computing Machinery, New York, NY, USA, pp. 668–671. http://dx.doi.org/10.1145/3501712.3538830.

Sweeney, C., Potts, C., Ennis, E., Bond, R., Mulvenna, M.D., O'neill, S., Malcolm, M., Kuosmanen, L., Kostenius, C., Vakaloudis, A., Mcconvey, G., Turkington, R., Hanna, D., Nieminen, H., Vartiainen, A.-K., Robertson, A., Mctear, M.F., 2021. Can chatbots help support a person's mental health? perceptions and views from mental healthcare professionals and experts. ACM Trans. Comput. Heal. 2 (3), http://dx.doi.org/10.1145/3453175.

Thieme, A., Belgrave, D., Doherty, G., 2020. Machine learning in mental health: A systematic review of the HCI literature to support the development of effective and implementable ML systems. ACM Trans. Comput.-Hum. Interact. 27 (5), http://dx.doi.org/10.1145/3398069.

Tong, X., Mauriello, M.L., Mora-Mendoza, M.A., Prabhu, N., Kim, J.P., Paredes Castro, P.E., 2023. Just do something: Comparing self-proposed and machine-recommended stress interventions among online workers with home sweet office. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. CHI '23, Association for Computing Machinery, New York, NY, USA, http://dx.doi.org/10.1145/3544548.3581319.

Valentine, L., D'Alfonso, S., Lederman, R., 2022. Recommender systems for mental health apps: advantages and ethical challenges. AI Soc. 38 (4), 1627–1638. http://dx.doi.org/10.1007/s00146-021-01322-w.

Victorelli, E.Z., Dos Reis, J.C., Hornung, H., Prado, A.B., 2020. Understanding human-data interaction: Literature review and recommendations for design. Int. J. Hum.-Comput. Stud. 134, 13–32. http://dx.doi.org/10.1016/j.ijhcs.2019.09.004, URL https://www.sciencedirect.com/science/article/pii/S1071581919301193.

Vossen, W., Szymanski, M., Verbert, K., 2024. The effect of personalizing a psychotherapy conversational agent on therapeutic bond and usage intentions. In: Proceedings of the 29th International Conference on Intelligent User Interfaces. IUI '24, Association for Computing Machinery, New York, NY, USA, pp. 761–771. http://dx.doi.org/10.1145/3640543.3645195.

Wang, Y.-C., Lu, Y., Grunwald, S., Chu, S.L., Kamble, P., Kumar, J., 2024. An AI approach to support student mental health: Case of developing an AI-powered web-platform with nature-based mindfulness. J. Hosp. Tour. Educ. 1–14.

Wu, R., Yu, C., Pan, X., Liu, Y., Zhang, N., Fu, Y., Wang, Y., Zheng, Z., Chen, L., Jiang, Q., Xu, X., Shi, Y., 2024. MindShift: Leveraging large language models for mental-states-based problematic smartphone use intervention. In: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems. CHI '24, Association for Computing Machinery, New York, NY, USA, http://dx.doi.org/10.1145/3613904.3642790.

Xygkou, A., Siriaraya, P., Covaci, A., Prigerson, H.G., Neimeyer, R., Ang, C.S., She, W.-J., 2023. The "conversation" about loss: Understanding how chatbot technology was used in supporting people in grief.. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. CHI '23, Association for Computing Machinery, New York, NY, USA, http://dx.doi.org/10.1145/3544548.3581154.

Yang, W., Chuah, M.C., Du, L., 2020. M&m: Deep learning aided multi-facet mental health support tool for college students. In: Proceedings of Deep Learning for Wellbeing Applications Leveraging Mobile Devices and Edge Computing. HealthDL '20, Association for Computing Machinery, New York, NY, USA, pp. 10–15. http://dx.doi.org/10.1145/3396868.3400898.

Young, J., Jawara, L.M., Nguyen, D.N., Daly, B., Huh-Yoo, J., Razi, A., 2024. The role of AI in peer support for Young people: A study of preferences for human- and AI-generated responses. In: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems. CHI '24, Association for Computing Machinery, New York, NY, USA, http://dx.doi.org/10.1145/3613904.3642574.

Zhai, X., Chu, X., Chai, C.S., Jong, M.S.Y., Istenic, A., Spector, M., Liu, J.-B., Yuan, J., Li, Y., 2021. A review of artificial intelligence (AI) in education from 2010 to 2020. Complexity 2021 (1), 8812542.

Zhao, Y., Li, T., Sobolev, M., 2024. Digital wellbeing redefined: Toward user-centric approach for positive social media engagement. In: Proceedings of the IEEE/ACM 11th International Conference on Mobile Software Engineering and Systems. MOBILESoft '24, Association for Computing Machinery, New York, NY, USA, pp. 95–98. http://dx.doi.org/10.1145/3647632.3651392.