

## Review

# Intelligent Cognitive Assistants for Attitude and Behavior Change Support in Mental Health: State-of-the-Art Technical Review

Tine Kolenik <sup>1,2,\*</sup>  and Matjaž Gams <sup>1</sup> <sup>1</sup> Department for Intelligent Systems, Jožef Stefan Institute, 1000 Ljubljana, Slovenia; matjaz.gams@ijs.si<sup>2</sup> Jožef Stefan International Postgraduate School, 1000 Ljubljana, Slovenia

\* Correspondence: tine.kolenik@ijs.si

**Abstract:** Intelligent cognitive assistant (ICA) technology is used in various domains to emulate human behavior expressed through synchronous communication, especially written conversation. Due to their ability to use individually tailored natural language, they present a powerful vessel to support attitude and behavior change. Behavior change support systems are emerging as a crucial tool in digital mental health services, and ICAs exceed in effective support, especially for stress, anxiety and depression (SAD), where ICAs guide people's thought processes and actions by analyzing their affective and cognitive phenomena. Currently, there is no comprehensive review of such ICAs from a technical standpoint, and existing work is conducted exclusively from a psychological or medical perspective. This technical state-of-the-art review tried to discern and systematize current technological approaches and trends as well as detail the highly interdisciplinary landscape of intersections between ICAs, attitude and behavior change, and mental health, focusing on text-based ICAs for SAD. Ten papers with systems, fitting our criteria, were selected. The systems varied significantly in their approaches, with the most successful opting for comprehensive user models, classification-based assessment, personalized intervention, and dialogue tree conversational models.

**Keywords:** artificial intelligence; behavior change support systems; digital mental health; e-health and well-being; intelligent cognitive assistant; technical state-of-the-art review



**Citation:** Kolenik, T.; Gams, M. Intelligent Cognitive Assistants for Attitude and Behavior Change Support in Mental Health: State-of-the-Art Technical Review. *Electronics* **2021**, *10*, 1250. <https://doi.org/10.3390/electronics10111250>

Academic Editors: Maria Goreti Carvalho Marreiros, João Carneiro and Jaume Jordan

Received: 30 April 2021

Accepted: 20 May 2021

Published: 24 May 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Change—what it is, why it happens and how to achieve it, especially in the human psyche—has taken many forms throughout recorded history: for the ancient Greek philosopher Heraclitus, change came from cosmic fire [1]; for the great Chinese teacher Confucius, change was perpetually created by the eternal struggle between opposing forces [2]; for the medieval theologian Thomas Aquinas, change originated from another world [3]; for the systems theory and family therapy pioneer Paul Watzlawick, change emerged from paradoxes [4]; for the information-ager, change is driven by technology. Strangely enough, all of them would be correct if referring to human psyche: Heraclitus' cosmic fire is love (commonly mythologized as one of the sources of change, e.g., by the pre-Socratic philosopher Empedocles [5]), presumed to be the seed that produces and sustains change [6]; Confucius' opposing forces represent cognitive dissonance, behavior opposing attitudes and beliefs, resulting “in a psychologically uncomfortable state that motivates people to reduce the dissonance [...] by changing their attitudes to be more consonant” [7] (p. 1469); Aquinas' another world is the world of the human mind; Watzlawick's paradoxes demystify Confucius' divine opposing forces through a pragmatic psychotherapeutic framework; and the most de nos jours of all—the information ager's notion of how technology influences us is currently one of the most broadly discussed topics that strongly conforms to the reality of living in the information society [8].

The marriage between the advances of behavioral sciences, especially the vast knowledge on the psychological theories of how to effect attitude and behavior change in people,

and technology, which has seeped its way into an omnipresent fact of our lives, has seemingly more than ever given us the possibility and the tools to do what has always been the holy grail of human endeavors—to produce effective and rapid change. The information society [8] is equipping us with intelligent informational sources at every step—a smartphone is always in our pockets, a smart bracelet always on our wrists, an internet connection in every nook and cranny of our paths. This gives us the possibility to not only always know how we are behaving, but also to intervene with that behavior.

Such pervasive technology can produce attitude and behavior change in many domains of human life where it is sought—we generally want ourselves and others to be healthy, which means exercising, sleeping and eating well; to feel well, which means successfully navigating situations and thoughts that lead us down the path of mental issues, such as stress, anxiety and depression (SAD); and other behaviors, of which most can be found in the UN's Sustainable Development Goals [9]. Issues, addressed by such behaviors, have in the recent decades seen a catastrophic rise. One issue particularly affected by the recent COVID-19 pandemic is mental health. The lack of resources and effective systemic frameworks in the field of mental is not a recent development, but it was the pandemic that had exposed how disastrous neglecting people's well-being for decades can be [10]. Existing systems were further incapacitated by imposed social distancing, where the bond between people and mental health experts severed drastically. Decision-makers are consequently turning towards technology to help in what is not only a pandemic of the body, but also a pandemic of the mind. This work contributes a piece in the needed mosaic of a systematized effort to identify how technology can help in tackling this mental health crisis. This section continues with a statement on our motivation for this work and why we believe it is needed. Afterwards, it presents three interconnected areas of research that meet for such interdisciplinary efforts: attitude and behavior change (ABC) support systems, intelligent cognitive assistant (ICA) technology, and digital mental health. Afterwards, it presents related review papers, highlighting why their insufficiency for our purposes and for computer science researchers in general, and ends the section with an outline of the paper.

### *1.1. Motivation for This Work*

The need for this work—a review paper on intelligent cognitive assistants for attitude and behavior change in mental health, specifically for stress, anxiety and depression, from a perspective of researchers in technology-related fields—arose from investigating the technological trends and underpinnings of dialogically driven technology, used for psychological help. The search turned out three kinds of papers: (1) review papers, written by clinical experts and psychologists for clinical experts and psychologists [11–16]; (2) experiments on symptom relief in people with mental health issues, where an ICA was used; and (3) scattered papers by researchers who designed their own ICAs for mental health.

Papers described in (1) (see the first paragraph of this section), and more thoroughly detailed in Section 1.5, although insightful and helpful in several ways, were not meant for experts in our field of research, as some of them clearly explicate: “This review aimed to inform health professionals” [11] (p. 11) and “[a]lthough [embodied conversational agent (ECA)] research is almost inherently interdisciplinary, we refrained from going too deep into the technological aspects. This was because our target audience consisted of health professionals with a generally less technical background and we wanted to focus on opening up the ECA domain for them as well as providing them with an overview of the available evidence for application in routine clinical practice” [11] (p. 13). They covered very little technical, if any, aspects of ICAs used in experiments to offer psychological help, and mostly focused on the effects they had on participants. Introducing technical aspects into such work was the first incentive for our present work.

The review papers that do exist on this topic, although not technical, covered a number of papers that presumably should offer some information on ICAs for mental health, used

in their experiments. However, such papers, described in (2) (see the first paragraph of this section), offer either little or no technical description of the systems used, or the systems used were proprietary and their technical details are not disclosed [17–19]. This made it clear that the works this review strives to examine was not included in previous reviews.

Papers in (3) (see the first paragraph of this section) are the kind of papers this work focuses on. To be rigid in our systematization, we applied various criteria that papers had to meet to be included in this review (for more, see Section 2.5). We mostly focused on work that was of technical nature and where it was clear from the system design which technologies and methods were used to provide effective psychological help and cause attitude or behavior change.

The need for the proposed review was identified before: “[I]t has to be noted that, depending on how one would like to use ECAs in future work, many more detailed questions could be investigated surrounding ECA design aspects, such as the required capabilities for, and their impact on, specific disorders or types of ECA interventions” [11] (p. 13). However, to the best of our knowledge, the existing review papers fall under the group of papers described in (1). The specific details that we are interested in as opposed to the latter, warranting our work, can be found in our Research Questions subsection.

This paper is the first systematized overview of intelligent cognitive assistants for attitude and behavior change in mental health with a focus on stress, anxiety and depression from a technical perspective. Due to the technical focus, we opted for the state-of-the-art review (for more, see Section 2.1). This novel research is necessarily interdisciplinary, combining knowledge from computer science, psychology, behavioral sciences, cognitive science, psychotherapy, and related fields. Multiple perspectives, drawing from the authors’ backgrounds, are offered on the topic, mainly from computer and cognitive science.

## 1.2. Attitude and Behavior Change Support Systems

ABC support systems are computer systems that attempt to “change attitudes or behaviors or both (without using coercion or deception)” [20] (p. 20) and to “aid and motivate people to adopt behaviors that are beneficial to them and their community while avoiding harmful ones” [21] (p. 66). Attitude and behavior change signifies a phenomenon that is considered to be a temporary or lasting effect on an individual regarding their attitude or behavior as compared to what their attitudes were or how they behaved in the past [21] (p. 66). ABC support systems belong to the larger family of persuasive technology (PT). PT is the result of the vast advances in behavioral sciences in regards to psychological change [22], human decision-making [23] and related phenomena [24] as well as the arrival of digital technologies, artificial intelligence (AI) and big data. Many societal efforts have been put into creating technologies that would help, motivate, guide and persuade people into bettering themselves and the world around them, though such technology can be and has been abused as well [25]). ABC support systems have also been in the forefront of research (e.g., at the world’s biggest AI conference, IJCAI) for helping achieve the United Nations Sustainable Development Goals, which include ensuring “healthy lives and promote well-being for all at all ages”, “inclusive and quality education for all and promote lifelong learning”, taking “urgent action to combat climate change and its impacts” and others [26]. PT is already used in the health and wellness areas, where it tracks people’s behavior as well as their physiological and psychological processes, responding to them by trying to affect their mental states by offering psychotherapeutic advice or to motivate them into making different decisions, e.g., in regards to healthy eating [21]. There are also applications in areas such as education or environmental sustainability, where people are nudged towards greener behavior [27].

Major persuasive and ABC frameworks [28], that such technologies employ, include: Cialdini’s Principles of Persuasion (CPP) [22], Fogg Behavior Model (FBM) [20] and Persuasive System Design Model (PSDM) [29], with firm verification of their effectiveness [30].

CPP is based on the idea that general persuasive strategies are not equally effective for everyone. It identifies various strategies that affect different groups of people differently.

Interactive, adaptive technology can be utilized to personalize itself to specific strategies that work for specific groups of people.

FBM is based on the idea that a certain behavior is the result of motivation, ability and a trigger occurring at the same time. Therefore, a person changing their behavior has to be sufficiently motivated, has to possess the ability to change the behavior, and has to be triggered to change the behavior. These are then combined in personalized ways to find the most effective strategies for an individual.

PSDM is based on the need for effective design and evaluation of persuasive systems, and mostly offers a framework for what kind of content and functionality PT should consider. PSDM includes four principles upon which to design PT: (1) primary task support, which supports the user's carrying out of their primary task; (2) dialogue support, which helps users move towards their goals; (3) system credibility, which raises the user's belief in the system's quality; and (4) social support, which motivates the user by leveraging social influence.

Another powerful and effective behavioral change concept—Richard Thaler, its author, received the Nobel Prize for it—is the 'nudge theory'. Nudge is "any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentive", where "the intervention must be easy and cheap to avoid" [24] (p. 6). Nudges are being incorporated into PT and ABC support systems as well [31].

For persuasive strategies to be as effective as possible, they have to be tailored to a number of specifics. There are 4 factors in the framework of the Communication-Persuasion Paradigm [32] that determine the influence: (1) characteristics of the source (i.e., the message sender); (2) the message; (3) characteristics of the destination (or the receiver of the message); and (4) the context.

For determining effective strategies, personality models, such as Big Five personality traits (B5) [33] or Hexaco [34], as well as domain specific questionnaires, offer PT a useful way to model a person. Personality is measured on different dimensions (e.g., in B5: openness, conscientiousness, extroversion, agreeableness, neuroticism), which try to describe psychological and cognitive functionalities of individuals, e.g., their mental states and decision-making abilities. Knowledge in specific domains relies on PT's use of questionnaires. For mental health, SAD questionnaires [35] can be used to categorize people with SAD symptoms, which leads to better strategy selection. Such questionnaires give insight into what influences which individuals the most. Empirical phenomenology can also be employed for more detailed first-person accounts [36], which can be used for extracting linguistic features [37,38] or for other tweaking of ABC techniques in PT. Furthermore, combining subjective data with physiological data is also proving useful for adaptive technologies [39].

These frameworks and models appear in several technological platforms. The most frequently used platforms are mobile and handheld devices (28%), followed by games (17%), web and social networks (14%), other specialized devices (13%), desktop applications (12%), sensors and wearable devices (9%), and ambient and public displays (5%) [21].

ABC can be delivered through various software systems. Intelligent cognitive assistants (or chatbot, chatterbot, interactive agent, conversational AI, smartbot, bot) seem to be the most advanced [11–16,40]. The next subsection introduces such systems and describes why they seem to be the best vessel for delivering ABC.

### 1.3. Intelligent Cognitive Assistant Technology

There is a lack of consensus regarding the term with which to label technologies this review describes. Conversational agents, dialogue systems, smart conversational interfaces, relational agents, chatbots, and so on—in the end, we decided to go into the direction of the SRC workshop on the ICA technology [41], and label it intelligent cognitive assistant (ICA) technology, as it seems to better describe the capabilities such systems are designed to have. They are not only intelligent in terms of being able to converse and

have a language model, they have many other abilities that are human-like, relating to human cognition and intelligence. ICA technology has therefore been touted as the next revolution in human–computer coexistence. The technology dates back to the beginning of AI, where one of the first chatbots was developed and available outside of a research laboratory—Weizenbaum’s simulation of a Rogerian psychotherapist called ELIZA [42]. However, technological progress has only recently laid the foundations for broad adoption in the form of ICAs such as Alexa and Siri as well as more domain-specific agents such as Woebot [17]. Alexa, Siri and Google Home, however close to certain human capabilities they may seem, still often fail outside of very basic, secretary-like tasks. When used in more expert domains, such as mental health, they quickly start repeating themselves, as they only have very generic models that end up in common phrases and trivial platitudes. Sometimes, their remarks can be even dangerous for the user, as they may be perceived as flippant and negative, or give wrong medical advice. Testing their response to stressful accounts, they either do not understand or they fail to show empathy beyond empty words [43]. Expert domains of engagement therefore need domain-specific ICAs.

ICAs, which can be deployed in many devices, e.g., as virtual agents or robots, are striving to: understand context; be adaptive and flexible; learn and develop; be autonomous; be communicative, collaborative and social; be interactive and personalized; be anticipatory and predictive; perceive; act; have internal goals and motivation; interpret; and reason. To be able to come close to such capabilities, ICAs are embedded with a cognitive architecture (CogA), a “hypothesis about the fixed structures that provide a mind, whether in natural or artificial systems, and how they work together—in conjunction with knowledge and skills embodied within the architecture—to yield intelligent behavior in a diversity of complex environments” [44] (para. 2). Most importantly, ICAs possess the ability to converse in natural language. This seems to be the most immediate way in which humans communicate [45], and the effects of a dialogue on human mental states cannot be overestimated. ICAs, coupled with ABC capabilities, are establishing as a very promising PT.

Using ICAs for ABC is still a new field of research, despite ELIZA being the first chatbot, as chatbots have mostly been explored for education, customer support or in other simple question-answer contexts [46]. What makes ICAs for ABC unique, is that users reveal personal information more freely, which makes systems more successful in their goals [47]. ABC ICAs and their users also form a more longitudinal relationship. The interactions are not a one-off, where it is difficult to understand the users and act immediately with efficient strategies. This makes such ICAs able to learn from historical interactions and improve in achieving ABC. However, there is a considerable lack of evaluation standardization of PT and ABC support systems, which makes the research field prone to the introduction of researcher bias.

ICAs, besides being a vessel to understand users through modeling their psychological and physiological aspects and use such knowledge to enact ABC, present as an ideal platform for offering help in the field of mental health because of the ability to converse. This opens up new solutions in the field of digital mental health.

#### 1.4. Digital Mental Health

Although the COVID-19 pandemic has revealed in full the problems that mental healthcare has [10] as well as thoroughly exacerbated existing well-being of people [48], the mental health pandemic has been raging for far longer [49,50]. Various decision-makers—especially world organizations, national governments and other leaders—are starting to recognize this, which is why mental well-being appears in Goal 3 of the 17 UN Sustainable Development Goals [26]. Most common mental health issues include stress, anxiety and depression (SAD); these have seen the biggest rise in the recent decades [51].

Before the COVID-19 pandemic, figures for SAD symptoms in some groups reached over 70% for overwhelming stress symptoms, which make people unable to cope [52,53] and about 8% for disorders, connected with stress, such as post-traumatic stress disorder.



der [54]; almost up to 34% for anxiety disorder [55]; and up to 27% for depressive symptoms [56] and 6% for depressive disorders [57]. With the COVID-19 pandemic, we are seeing these numbers rise [48,58,59]. The number of people with SAD symptoms and disorders increased up to more than three-fold [60]. What is even more worrying is that mental health issues are very underreported, especially in developing countries [61]. How different countries report the state of their population's mental health is also rough at best and the data are mostly about the adult population [55]. This is further skewed by the fact that up to 85% of people in low- and middle-income countries receive no mental health treatment [62], treatment coverage in high-income countries for certain disorders only reaches 33% [63], and up to 96% people with SAD do not seek treatment at all [64].

Mental health issues have substantial, multi-faceted consequences, not only affecting the patient, but also their immediate surroundings (family, caretakers) and the wider society [65]. Patients are faced with a decreased quality of life, poorer educational outcomes, lowered productivity and potential subsequent poverty, social problems, abuse vulnerabilities, and additional non-mental health problems. Patients' immediate family and caretakers face increased emotional and physical challenges, decreased household income, and increased financial costs. Society as a whole faces exacerbating public health issues, corrosion of social cohesion, and the loss of several GDP percentage points and billions of dollars expenditure per nation annually. What ends up happening is that SAD increasingly perpetuates SAD. Too often, the direct result of this is the worst possible one—loss of human life. Many countries still struggle with a high suicide rate [51]. The reasons for increasing of SAD include a critical lack of mental health professionals and regulations [66] as well as inequality in access to care [67,68].

The conditions in which mental healthcare finds itself in, especially in a post-COVID-19 world, seem to present an opportunity for development of technological and other scientific therapy-based interventions, especially as individuals with mental health issues prefer therapies to medication [69]. Digital mental health, a still insufficiently explored area of research and practice, represents a way to explore how technology can complement existing mental healthcare systems to be more effective in delivering help to people that need it.

Technologies that increase the operability and effectiveness of healthcare are many [70–73], but we concentrate on addressing the implications using ICAs as PT in mental health has. By not only focusing on what ICAs offer but also on possible problems they bring, we try to provide a fair account of the potential of digital mental health as a whole.

We identified the following areas where using PT can offer positive possibilities in mental healthcare: cost, availability, stigma, and prevention. Identified negative possibilities include group exclusion, research bias, privacy problems, lack of longitudinal research, ethics of using personal information for persuasion, potential risks of digital dependence, potential problems of automation and job loss of mental healthcare professionals, and possible cost increase in certain aspects.

#### *Positive possibilities:*

Cost related to the service of mental healthcare professionals varies, not only due to country standards, but also on country regulations and subsidies. It highly depends on the number of practicing professionals. Regardless, it presents a barrier for people of lower socio-economic backgrounds [74]. PT for mental health can be realistically made free of charge (and many times is [17]) due to the much lower costs attached to it [68]. PT also offers collecting data on often overlooked (and disadvantaged) populations, thus lowering systemic bias in analysis, as well as targeting patients with low-priority conditions.

Availability refers to location, time, and cost. Location-based availability concerns people with no direct access to mental healthcare (e.g., remote areas) [75]. Time-based availability concerns people needing help when their chosen professional is unavailable (e.g., panic attack during the night). PT may also minimize problems related to transportation [76]. Cost-based availability concerns people needing more than the minimum recommended

amount of hours of psychological help per week [77]. Research [77,78] shows that more frequent therapy results in better outcomes, and complementary use of PT for mental health can bridge that gap for people not being able to afford more therapy by still having access to help.

Stigma refers self-stigma—the prejudice which people with mental issues turn against themselves—and public stigma—the reaction that the general population has to people with mental issues. Both are prevailing problems [79], causing up to 96% people not deciding to seek treatment [64]. Research shows that people are more comfortable disclosing themselves to a computerized system than to a person [47]. Introducing PT for this group of people may result in offering help to people that would otherwise never receive it as well as helping people get better to the point of visiting a professional.

Prevention refers to the blind spot in mental healthcare: people only come in (if at all) seeking treatment, while a lot of issues can be prevented beforehand. PT can work indirectly by providing “support for better decision making, emotional regulation or interpersonal interactions,” which are “necessary to ensure that psychological, emotional and social deficits do not spiral into clinical disorder,” or directly by improving “both the screening and early delivery of interventions to reduce risk factors and build psychological resources” [76] (p. 336). Therefore, treatment should not be the only target for PT, as prevention also lowers the amount of mental health issues present and thus relieve stress on healthcare.

#### *Negative possibilities:*

Group exclusion refers to those groups of people which can not only be excluded from technology-oriented mental healthcare, but may find themselves even further distanced from or even removed from the society. The groups include the elderly, who have difficulties integrating technology into their lives [80]; the lowest socio-economic class, who may not benefit from PT due to their lack of access to technology [81]; and culturally-specific groups, whose cultural or sociopolitical specifics prevent them from adopting technology [82]. Fortunately, PT research is fledging in certain low-income parts of the world [83].

Researcher bias refers to the lack of evaluation standards of PT for mental health in this interdisciplinary endeavor. This is due to two factors: the field’s youth and the various disciplines tackling the field individually. The possible problems are many: (1) PT are not always studied in empirical experiments, but in quasi-experiments [84] or no experiments at all, but if there are empirical experiments, it is mostly with PT that is proprietary and is thus harder to change; (2) the metric on which to evaluate such systems is unclear (usually comes indirectly from their effectiveness in an experiment where the goal is SAD symptoms relief [12]); (3) no consensus on what data is needed. This results in many unfounded presuppositions of researchers, ending in problematic practice.

Cost increase refers to the possibility that using PT in mental health may delay “the provision of traditional treatments with greater evidence of efficacy or by increasing the numbers of patients receiving services” [76] (p. 336). More research is needed to be able to understand the costs alleviated and costs incurred by implementing such systems.

Other potential problems are less related to our work, but worth the mention nevertheless: (1) the problem of personal information privacy [85]; (2) the problem of the lack of longitudinal research on behavior change with PT [86]; (3) the ethics of using personal information for persuasion [85]; (4) the potential risks of digital dependence [87]; and (5) the potential problem of automation and job loss of mental healthcare professionals.

Reviews on this topic are favorable [11–16,40], agreeing that “early evidence shows that with the proper approach and research, the mental health field could use conversational agents in psychiatric treatment.” [12] (p. 456). Related review papers are presented next.

#### *1.5. Related Work*

Due to the novel viewpoint of our review, deciding on the parameters of what constitutes as related work was non-trivial. It was established that none of the found review

papers covered ICAs that try to induce ABC for SAD from a technical point of view—analyzing their software structures, algorithms used, datasets which they utilize, etc. The review papers we present in this section therefore consists of work that analyzed such systems in a way that is “aimed to inform health professionals” [11] (p. 11) as opposed to researchers in the fields of computer science.

Related work is divided into three groups: (1) papers that review the use of ICAs (under different synonyms—conversational agents, chatbots, etc.) for delivering help in mental health; (2) papers that review the use of applications in general for delivering help in mental health; and (3) papers that review the use of ICAs for delivering help in health in general.

We identified six related works, belonging to the first group of papers. Provoost et al. [11] focused on embodied conversational agents (ECAs), which beside language also simulate some properties of human face-to-face conversation, including non-verbal behavior. They tried to provide an overview of the possibilities such systems present and to investigate the evidence base for their effectiveness. They found 54 studies with ECAs for treating mood disorders, anxiety, psychoses, autism, and disorders connected to substance use, which use different techniques, including reinforcement of social behaviors through expressions and multimodal conversations, to reduce symptoms. They concluded that this avenue presented an emerging and important research endeavor, with the limited results so far showing positive outcomes. They also called for more research and the production of more such systems. Vaidyam et al. [12] explored chatbots in psychiatry for assessment as well as intervention purposes. They focused on chatbots for depression, anxiety, schizophrenia, bipolar and substance abuse disorders. From 10 studies that fit their criteria, they found that the reported outcomes in using chatbots showed benefit in psychoeducation and self-adherence, as well as it being an enjoyable tool that patients used. They concluded that early evidence was promising, however they called for more research from all the actors in this interdisciplinary field. Abd-Alrazaq et al. [13] identified chatbots as a possible remedy for the shortage of mental health workers, which prompted them to pool effectiveness and safety results of 12 studies on using chatbots for depression, distress, stress, and acrophobia. They found that there is a lack of evidence on whether their effect was clinically important, but they concluded that they are safe. They warned that there is a lack of standardized evaluation metrics, resulting in high risk of bias. Gaffney et al. [14] investigated ECAs and their usability for general psychological distress. In 13 identified studies, they discovered that the efficacy and acceptability were promising with most studies showing significant reductions in mental issue symptoms. They called for researchers to produce more work on exploring mechanisms of change such systems can employ to increase efficacy, be it technical or not. Bendig et al. [15] focused on chatbots used in clinical psychology and psychotherapy. They found that most experiments done are pilot studies where it is hard to produce high-quality evidence. They report that practicability, feasibility, and acceptance of chatbots was very promising, although such technologies were still highly experimental, especially because applying technology in such a complex domain is difficult. They ended the review calling for funding to evaluate chatbots on effectiveness, sustainability, and safety. Abd-alrazaq et al. [16] reviewed chatbots for mental health, not excluding any mental disorders or purposes of chatbots in mental health. They found 41 chatbots, some only used for screening ( $n = 10$ ) or training ( $n = 12$ ), while other were used for therapy ( $n = 17$ ) or without a specific purpose. Most treated depression ( $n = 16$ ) or autism ( $n = 10$ ). As the authors before, Abd-alrazaq et al. called for more evidence, but recognized possible utility of early integration of such systems in mental healthcare.

We identified four related works, belonging to the second group of papers. Bakker et al. [88] focused on any mental health app for mental health. They discovered that they lack functionality and features. They also noted a lack of research on the efficacy of apps, worrying about a complete lack of trials of any kind. They presented their own recommendations for developers of such apps. Orji and Moffatt [21] reviewed persuasive technology from the span of 16 years and they comprehensively detailed their designs, research methods,



strategies and theories they use to persuade, as well as targeted behaviors. They concluded that persuasive technology was a promising avenue for wellness, but that the field was lacking longitudinal research and current technological limitations. Chan et al. [89] surveyed the use of mobile apps in psychiatric treatments. They called on mental health practitioners to show a bigger understanding for using such apps, what their features were, what should be studied more to advance their capabilities and what the possible issues may be in integrating them into clinical workflows. They concluded that patients with various mental illnesses and severities may benefit from them, despite their social and technological backgrounds, however, better practices for evaluating apps, understanding user needs, and educating them on their use was needed to increase the apps' efficacy, on top of ensuring ethical and risk-free protocols. Torous et al. [90] researched smartphone apps and focused on their adoption by clinics or consumers, as the uptake was still low in spite of the potential of apps to improve quality and access to mental healthcare. They reported high heterogeneity in metric reported by studies, and found that despite apps being even successful in their goals, they lacked user testing, privacy protection, and mechanisms that establish trustworthiness. They also did not tackle emergencies. They called for further research in all fields connected to this technology.

Finally, we identified five related works, belonging to the third group of papers. Laranjo et al. [40] focused on conversational agents with unconstrained natural language input capabilities for any health-related purpose, targeting customers as well as health professionals. They analyzed 14 different conversational agents with mostly finite-state or frame-based dialogue systems, focusing on patient self-care. Very few presented non-quasi-experimental studies. However, most reported satisfying efficacy, but they rarely evaluated patient safety. Authors ended the paper calling for better experimental designs and standardization in such works. Montenegro et al. [91] developed a taxonomy based on 40 papers related to conversational agents applied to healthcare, and with the taxonomy identified existing challenges and research gaps. They found many systems supporting patients as well as physicians, with a minority of systems focusing on student training. Most of the agents surveyed focus on health literacy, which the authors considered a future trend in the future of changing health behaviors. They discovered that the most lacking areas were bringing such technology to the elderly and making advances in user involvement, which included better interactions, interfaces, and models of learning. Safi et al. [92] investigated chatbots in the medical field from a technical aspect pertaining to their development. They identified 45 studies on using chatbots for health purposes. The most common method was pattern matching method, used commonly for question-answer conversations in providing information users ask for. Generating original output, not a pre-existing one, was rare. Very few studies collected any user data. The authors found such systems useful for providing information to interested users. Abd-Alrazaq et al. [93] performed an overview of technical (non-clinical) metrics used for evaluating dialog agents in healthcare. By scanning 65 studies, they found 27 technical metrics, pertaining to chatbots generally, to their response generation and understanding, and to their aesthetics. Their work tries to systematize and push the direction towards standardization of how to evaluate chatbots non-clinically. Pereira and Díaz [94] surveyed chatbots for health behavior change. They identified 30 papers that used health chatbots in their study, and found out that nutritional disorders and neurological disorders were the most targeted health issues, that the chatbots tried to change human competence in tackling these issues, and that users most appreciated the personalization and consumability aspects of these chatbots. Again, the authors noted that case studies were lacking and that technological implications were almost never discussed.

The rest of the paper is organized as follows: Section 2 presents the materials and methods used for this review, focusing on research methodology, study design, research questions, search strategy, paper selection criteria, and data extraction. Section 3 presents the results, focusing on search results and paper selection, description of selected papers, main findings, and answers to the research question. In Section 4, the work is discussed

and compared to existing review, the technology is evaluated and advantages and disadvantages are listed. The paper finishes with Section 5.

## 2. Materials and Methods

### 2.1. Research Methodology

To achieve the goal of conducting the first technical review of ICAs for ABC in mental health, we opted for state-of-the-art (SOTA) review with some elements of scoping review. Initial exploration of the literature led us to the same conclusion as well, as it revealed that more traditional systematic reviews, which put more emphasis on clear outcomes, or meta-analytics approaches, which require more comparable outcomes, are inappropriate due to the novelty and technical aspects of the field. SOTA reviews are especially appropriate for more technical analyses, especially in fast-evolving fields of study. The review method is also appropriate when the work is more exploratory, when, as in this case, systematization of such technologies does not exist yet. What is considered SOTA in our topic of review is as of yet unclear. However, to really focus on applicable trends and directions of research, our review covers research from 2016 to 2021 (approx. 5 years), which is not an uncommon timespan for fast-developing fields [95]. We believe this limited timespan enables us to only survey the latest developments, methods and technologies used for ICAs in mental health. SOTA review therefore helps us underpin key concepts in a research area and produce a summarized content, offering a better overview than other forms of review methods, and yielding consistent results to solidify new technological phenomena. Using SOTA review instead of other types should also appropriately differentiate this review from the reviews listed in Section 1.5, which largely focus on clinical outcomes instead of technical foundations.

Key objectives of this review are therefore to present a novel technological research area and its technical trends.

### 2.2. Study Design

Our research process found the suitable framework in Arksey and O'Malley's framework [96] for reviewing work. The framework provides a direction for the necessary steps in the process. The course of such an approach includes: (1) identifying the research questions; (2) identifying relevant works; (3) identifying selection criteria and applying it to step (2); (4) extracting and organizing the data; and (5) reporting the results in ways to address the research questions and satisfy the purpose of the review.

All the steps were followed by the authors as recommended in various works [97,98]. Stage (1) in the framework was conducted with regular discussions between the authors; stage (2) was conducted by the authors working individually, relying on their experience in the field and resolving any consequent discrepancies mutually; stage (3) was based on the goals of this work and the experience of the authors; stages (4) and (5) were conducted with close cooperation between the authors. Considerable attempts were made to provide a transparent and clear presentation of the research work which resulted in this paper.

### 2.3. Research Questions

The work does not have one central, scoping question. Technical trends are generally reflected in a number of subsystems that comprise one system, which led us to a collection of specific questions mostly regarding such subsystems. The systematization follows this process instead of being embodied in the questions themselves. However, answering these questions should lead to the heart of the phenomenon all interdisciplinary actors in this field are interested in: How do reviewed systems achieve change?

Research questions (RQs):

**RQ1.** Which mental health issues do the systems target?

**RQ2.** Which technologies, methods and collected data guide the process to achieve ABC for SAD in the systems?

**RQ3.** What are the technical aspects of the conversational models in the systems?

**RQ4.** What are the platforms used to create the systems?

**RQ5.** What domain knowledge is used to achieve ABC for SAD?

**RQ6.** What user modeling, especially for personalization and adaptation, do the systems conduct?

**RQ7.** What is the overarching cognitive architecture used in the systems?

**RQ8.** How are the systems evaluated in terms of ABC for SAD?

Some RQs will not have clear answers, especially as some would need clear standardization or metrics, currently not present in the field, and each of these would warrant a paper of its own. For example, evaluation, technical and clinical, seems to be left to the researchers' own devices each time they do a study, without any guidelines from a wider community. This is also why we are rather focusing on trends than on poorly defined metrics. Our RQs try to provide some direction in terms of which technical aspects of these systems are important to computer science-adjacent researchers, which is why we define our scope through them.

#### 2.4. Search Strategy

The search query was constructed with the authors independently collecting keywords and correlating them with synonyms and related words. The construction was based on the authors' knowledge of the area as well as referring to the review papers described in Section 1.5. We also used the PICOC methodology [99] to further refine our search string. Below is the mutually agreed upon search query:

*"chatbot" OR "conversational agent" OR "relational agent" OR "virtual agent" OR "intelligent agent" OR "cognitive agent" AND "anxiety" OR "depression" OR "mental health" OR "stress"*

Preliminary searches on a wide range of databases were conducted, including querying Scopus, PubMed, EBSCOHost, Springer, the ACM Digital Library, IEEE Xplore, Google Scholar, Web of Science, EmBase, PsycINFO, Cochrane, CINAHL, Science Direct, and Inspec. However, we found that Google Scholar (as an aggregator of various scientific works as opposed to a specific database with limiting inclusion criteria) had a wide enough coverage to allow it be used instead of the listed databases. We decided for this after discovering that this insight is consistent with empirical studies on database comparison [100–102]. Therefore, we decided to use only Google Scholar and complementing it with a database search software Harzing's Publish or Perish, which is recommended for easier searching [103].

The search was started on 24 March 2021.

#### 2.5. Paper Selection Criteria

Relying on our experience and knowledge of the field as well as having a clear idea of related work, we constructed a list of special criteria to apply to the paper selection process. We tried to elaborate on every decision to avoid arbitrary or biased criteria. We included all the full papers that passed all the items on the special criteria list.

The special criteria include the following items:

1. **Targeted mental health issues in the paper include stress, anxiety, depression, or general well-being.** We opted for this criteria as these are the most common mental health issues among the nonclinical population [57], they are seeing the most rise [104], they are targeted most by the systems we are interested in (see Section 1.5), and, as ascertained from related works, they are the easiest to target with technology.
2. **The system in the paper is autonomous and not Wizard-of-Oz.** The Wizard-of-Oz technique refers to the "seemingly autonomous application whose unimplemented functions are actually simulated by a human operator, known as the Wizard of Oz" [105] (p. 7). Since we are investigating technologies that enable exactly such functions, including Wizard-of-Oz systems would defeat the purpose of our work.

3. **The conversational model of the system in the paper is text-based.** We opted for text-based systems due to experts calling for such systems [11], due to our belief that text-based systems are the most mature in the technological landscape and therefore more amenable to being reviewed, and due to the amount of such systems being too wide to cover with one paper as the technologies used for, e.g., speech-based systems means analyzing a completely different technology.
4. **The conversational model of the system in the paper allows for a synchronous, real-time two-way communication.** We opted for this criterion due to the power of a dialogue in the matters of mental health [106], which compels us to research such systems, as well as the trending usage of ICAs in various areas of service [107], whose success also stems from the convenience of synchronous communication, seen in instant messaging systems [108].
5. **The paper describing the system was published between 2016 and 2021.** We opted for a SOTA review of the field, and since technology is developing fast, the last five years, recommended by other researchers as well [95], should cover the trends we want to observe.
6. **The system in the paper is implemented to be used with a computer or mobile devices.** We opted for this criterion—as opposed to also covering, e.g., robotic platforms—due to wanting to overview conveniently available systems, which do not demand additional resources for being accessible.
7. **The system in the paper is fully functional, not a part of a bigger cognitive architecture of an ICA.** The power of ICAs lies in their emergent behavior when multiple parts or modules work in concert towards producing ABC. We are interested in the system as a whole, not individual parts.
8. **The system in the paper is not only a design, but was implemented and can be used.** We want to explore systems that are possible to build. Only implemented systems can answer some of our questions (e.g., RQ4), especially on results that such systems produce (e.g., RQ8). We therefore believe that without this criterion, the true technical trends of the field cannot be sufficiently addressed.
9. **The paper describes the system in enough technical details to be able to analyze it from a computer science perspective.** To be able to conduct a SOTA review, this criterion is necessary. Without it, barely any RQ can be addressed.
10. **The system in the paper is non-proprietary.** Many systems (or platforms used to build the systems) used in the most (cited) studies [17–19], reviewed by papers in Section 1.5, are non-proprietary. The most well-known systems or platforms are: Tess [18], Wysa [19], Woebot [17], DialogFlow [109], IBM Watson [110], Microsoft Bot Framework [111], and GPT-3 [112]. Unfortunately, proprietary work cannot be surveyed as their technologies are closed source and not described in enough detail to be able to analyze them. They work as a double black box—not only cannot we discern the neural networks they use for their conversational models, we cannot even discern other methodological and technological details about them. We also want to foster open source and transparent research work, so our focus on analyzable systems should also be seen in light of this.

Another criterion that we were seriously considering, but knew we could not include, was for the system in the paper to be open code and be publicly available. However, this is still such a rarity that we quickly dropped the idea.

Apart from the specific criteria list to apply to paper selection, we constructed general criteria, partly guided by the PICOC method [99].

The steps that we followed for paper selection were:

1. Use of Harzing's Publish or Perish for easier management
2. **Exclusion criteria:** Papers do not address "Conversational Agents" and related acronyms (population criterion I)
3. **Exclusion criteria:** Papers do not address "Stress," "Anxiety," "Depression," or similar words (intervention criterion II)

4. **Removal of impurities:** Deleting theses, dissertations, non-scientific papers, posters, review papers, books, papers with three pages or less in length
5. **Quality assessment:** Focusing only on peer-reviewed published papers in journals and conferences (conferences hold special importance in computer science)
6. **Abstract and text filtering:** Special selection criteria, not applied before, described under the special criteria subsection

Removal of duplicates was not strictly necessary due to the use of one database, but since there might be various sources for the same paper (e.g., a journal and a university website), they were removed in one of the steps (e.g., step 3) or by hand when encountered.

We used PICOC to refine our criteria to be transparent and unbiased for the final paper selection. We took inspiration from the PRISMA framework [113] for reporting and we used the PRISMA diagram to visualize the process.

### 2.6. Data Extraction

Data extraction was focused on identifying keywords and parts of the text that help answering the RQs. Both authors independently extracted data from the papers which they deemed relevant to the review's narrative and goals. Afterwards, they relied on mutual agreement for combining the extracted data.

## 3. Results

This section presents the outcome of the review process. We report the search results and the paper selection process, we shortly describe the papers from the final selection, present the main findings and describe how they answer our RQs.

### 3.1. Search Results and Paper Selection

The paper selection process used various filtering methods to improve the results that fit the objectives of this review and help us answer our RQs. The process included the following steps: using Harzing's Publish or Perish for easier management; ad hoc removal of duplicates; application of exclusion criteria; removal of impurities (deleting theses, dissertations, non-scientific papers, posters, review papers, books, papers with three pages or less in length), application of quality assessment criteria, and abstract and text filtering. All authors independently selected the papers and mutually agreed on the final selection.

The search and selection **ended on 26 March 2021**.

The paper selection process with the numbers of papers encountered in each step was:

**Step 1:** Querying Google scholar with search string:  $n = 14,300$

**Step 2:** Using Harzing's Publish or Perish, applying exclusion criteria (population criterion I and II):  $n = 254$

**Step 3:** Removal of impurities, quality assessment:  $n = 114$

**Step 4:** Filtering:  $n = 10$  (number in line with similar review papers)

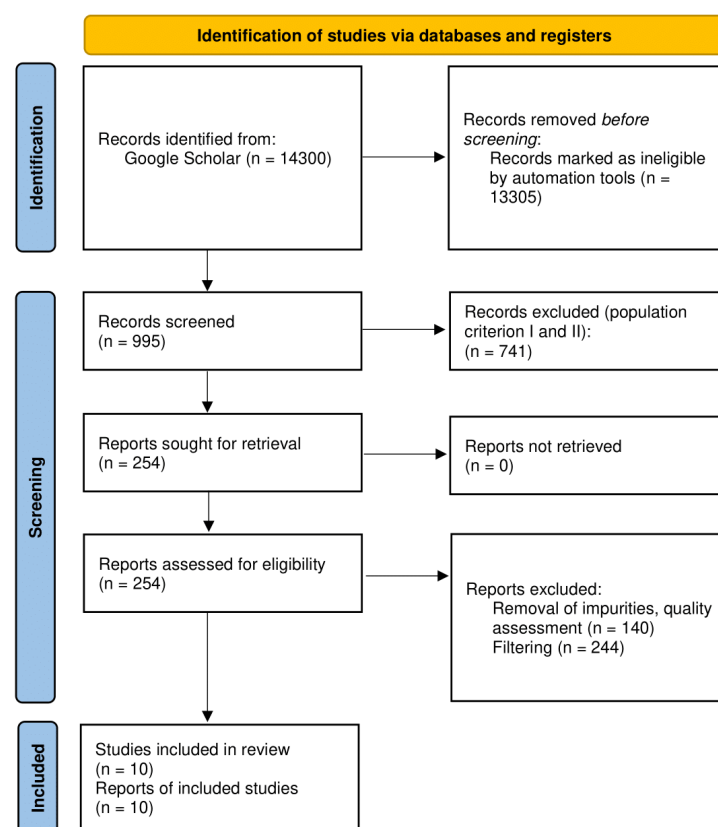
The PRISMA diagram in Figure 1 visualizes the process. The diagram follows the PRISMA methodology [113].

### 3.2. Description of Selected Papers

The selection process yielded 10 papers that aligned with our criteria. These papers represent various approaches to achieving change in people with mental health issues. Since all of them feature full cognitive architectures for their systems, some of the latter's parts are homogeneous among the papers, while others are very heterogeneous. Different means of achieving the same outcome is a much needed pluralism that new fields of research should always be adopting, especially when outcomes refer to SAD symptom relief in people with mental issues, which is an exceptionally complex process to undertake. The systems in this review show that there are multiple ways of doing that, which gives



the research field the flexibility and diversity. The two are needed for more possibilities for progress.



**Figure 1.** PRISMA diagram of the paper selection process.

Delahunty et al. [37] proposed a diagnostic ICA, which combined conversational abilities with machine learning and clinical psychology. It used sequence-to-sequence neural networks for dialogue generation and machine learning classifiers for discovering depression symptoms. The goal was to facilitate crisis support for depressive people.

Denecke et al. [38] introduced SERMO, an ICA that combined methods from cognitive behavior therapy (CBT) and lexicon-based emotion recognition to support general well-being in people by regulating their emotions, thoughts, and feelings. Emotion recognition in SERMO was crucial for effective strategy selection in terms of proposed activities and dialogue help. Alongside, informational strategies helped provide people with psychoeducation. User evaluation with the User Experience Questionnaire showed that the system was considered good.

Ghandeharioun et al. [114] focused on delivering ecological momentary interventions through an ICA to raise people's general well-being by relieving SAD symptoms. The system EMMA provided emotionally appropriate interventions in an empathetic manner, detecting user's moods solely through the smartphone sensor data, which was integrated with the ICA. Their results showed that their personalized machine learning model, used to determine the moods, was likable by the participants.

Khadikar et al. [115] developed Buddy, an ICA that targeted general well-being by treating symptoms of SAD, but also working as a motivational companion to help with loss of focus. The system used recurrent neural networks (RNNs) to respond to the users' emotions with appropriate dialogues that built mental resilience and drove the conversation towards positive thoughts.

Morris et al. [43] designed an ICA that simulated human capabilities in empathy expression. They repurposed online peer support data, which the ICA through corpus-based approaches presented to the user. Information retrieval and word embedding

techniques produced the best matches to the user's concerns. In a controlled experiment, the users found such responses acceptable.

Park et al. [116] delivered a prototype ICA Bonobot that used motivational interviewing methods to help students cope with stress. It used conversational sequences to guide the users through the motivational interviewing processes, providing evocative questions, encouraging feedback, and reflective and affirming responses, placed in the context of the users' problems. The major focus of Bonobot was discussing the idea of change. When used in an experiment, participants were satisfied with the ICA, but pointed out that more personalized feedback and informational support would benefit the system.

Pola and Chetty [117] created an ICA that offers behavioral therapy to people with depression. The ICA tried to get information from the user on their mental state. It could detect seven types of emotions from text using long-short-term-memory neural network and a pre-trained weighted word index known as glove2. The ICA's main strategy was trying to have a dialogue about the users' negative thoughts and offer different perspectives on them.

Rishabh and Anuradha [118] built three different ICAs for general well-being, using different technologies. The first, based on the famous psychotherapeutic chatbot ELIZA [42], used retrieval approaches for its language capabilities. The second, based on another famous chatbot, ALICE [119], used AIML (Artificial Intelligence Markup Language). The third used generative approaches. All of them tried to gauge the context that users conveyed to them through text and guide the conversation towards more positive sentiment.

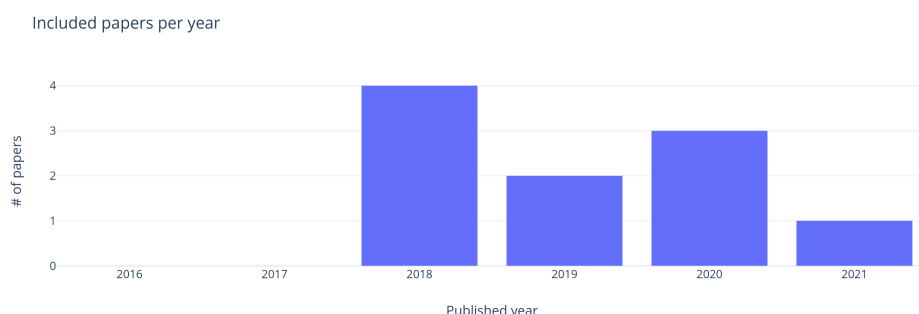
Yorita et al. [120] proposed a stress management framework with an ICA platform working on computers, mobile devices as well as in robots. It derived various stress measures and modeled their users, which determined the strategy selection in their peer support model. Interventions targeted various factors that aim at different stress management skills. The process was driven by reinforcement learning in combination with fuzzy control. Their results show that after using the ICA, people displayed better skills at dealing with stress.

Yorita et al. [84] built on the ICA from Yorita et al. [120], expanding the models and employed strategies for help to personalize their system even further.

### 3.3. Main Findings

This subsection presents some of the more general findings that led us to answer our RQs in the next subsection. The overall summary of our findings is presented in Figures 2–9, while more in-depth findings can be found throughout Section 3.4 and parts of Section 4.

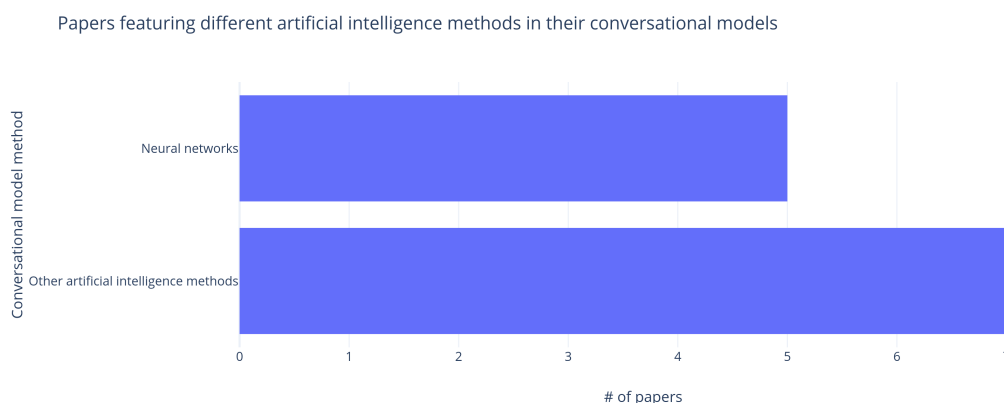
Figure 2 is the most general one and presents included papers per year.



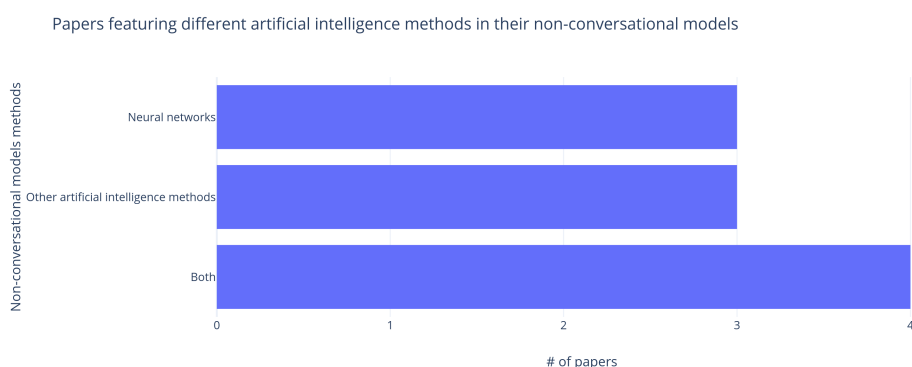
**Figure 2.** Included papers per year. No eligible works were found in 2016 and 2017.

Figures 3–6 represent the technical summary of the reviewed papers. Figure 3 shows the amount of papers that featured ICAs with conversational models based on neural networks, being the most popular generative method for natural language understanding

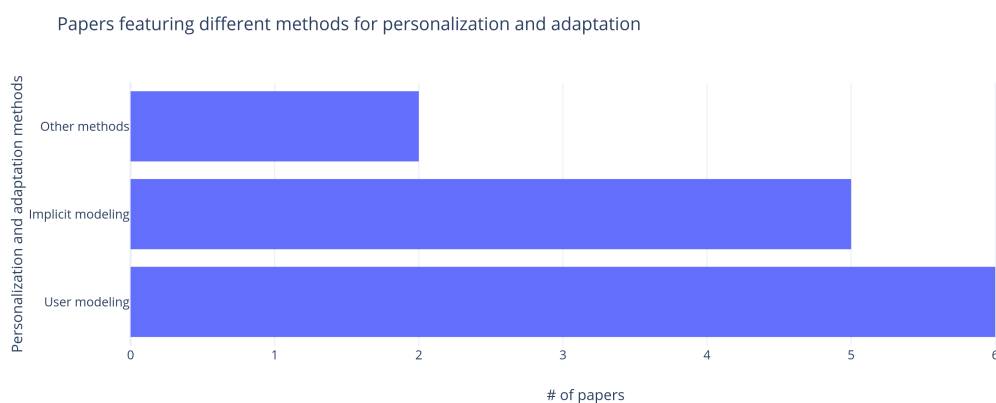
and generation, and the amount based on rule-based or other machine learning types. Figure 4 shows the amount of papers that featured ICAs with non-conversational models (e.g., classifiers for stress level) based on neural networks and the amount based on rule-based or other machine learning types. Figure 5 shows papers that featured ICAs that used various methods to personalize and adapt their actions. Figure 6 shows the amount of papers that featured ICAs that built their own complete cognitive architectures, and the amount that used existing (open source) platforms to create their architecture or that used existing ICAs and upgraded them.



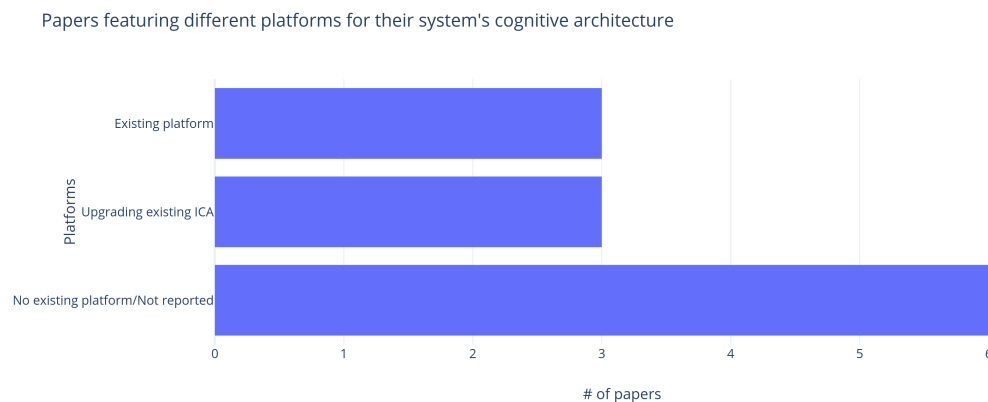
**Figure 3.** Papers featuring different AI methods in their conversational models. Some systems use neural networks as well as other AI methods, which puts them into both categories.



**Figure 4.** Papers featuring different AI methods in their non-conversational models.

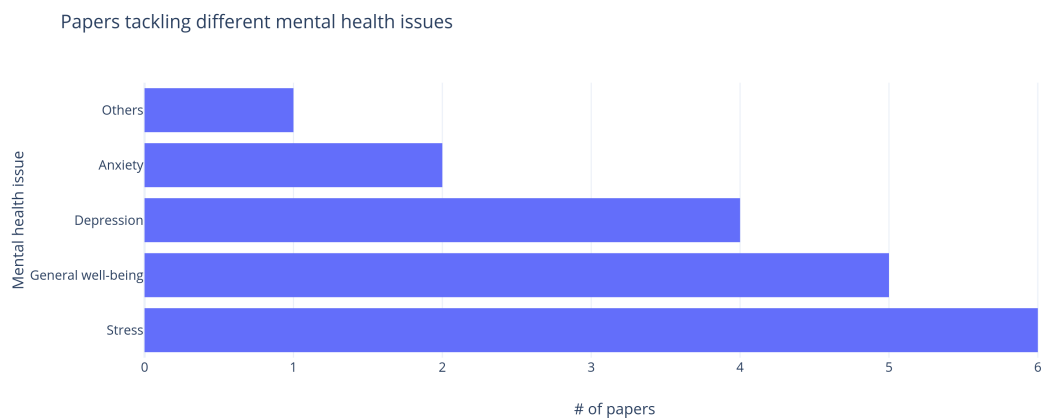


**Figure 5.** Papers featuring different methods for personalization and adaptation. Implicit modeling represents language understanding and generation methods, as ICAs personalize output by, e.g., recognizing emotions in the input.

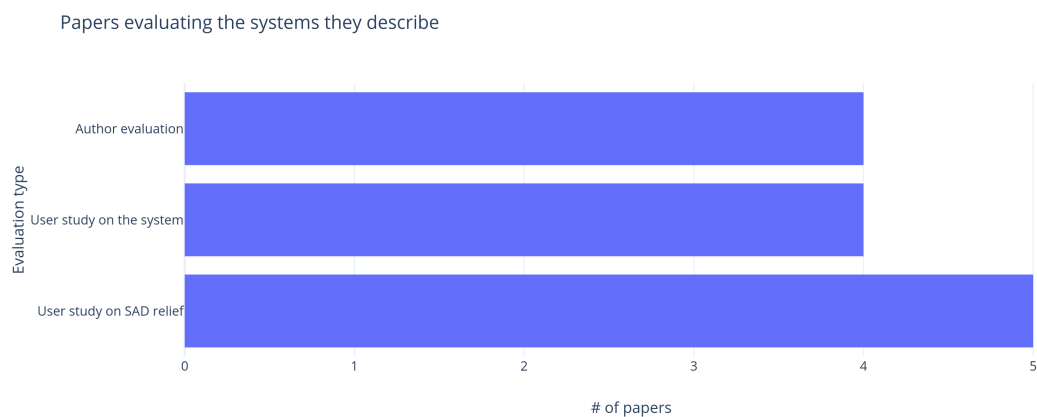


**Figure 6.** Papers featuring different platforms for their system's cognitive architecture. "Upgrading existing ICA" denotes using existing instances of architectures and upgrading them (e.g., ELIZA [116,118]).

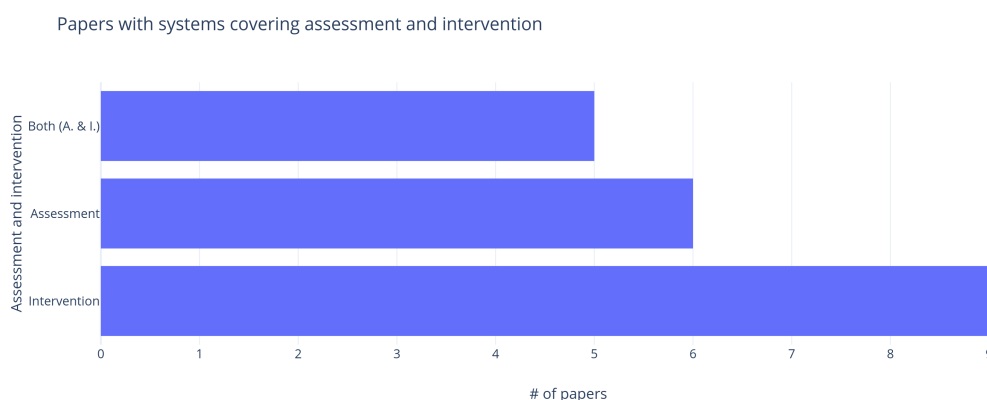
Figures 7–9 represent non-technical summary of the reviewed papers. Figure 7 shows the amount of papers that featured ICAs tackling specific mental health issues. Figure 8 shows the amount of papers that featured a user study on relieving SAD, that featured a user study on the system, and that were only evaluated by the authors. Figure 9 shows the amount of papers that featured ICAs that only did assessment, that only did intervention, and that did both.



**Figure 7.** Papers tackling different mental health issues.



**Figure 8.** Papers with different system evaluations.



**Figure 9.** Papers with systems covering assessment and intervention.

### 3.4. Answering the Research Questions

To answer the research questions, both authors independently identified parts of the reviewed papers, relevant for each RQ. The extracted information was synthesized by mutually agreeing on what information answers our RQs. To present the answers in a transparent and clear way to allow for easier comparison between the reviewed works, we answer each RQ with a comparison table.

#### 3.4.1. RQ1. Which Mental Health Issues Do the Systems Target?

To answer RQ1, we scanned the reviewed works for information on which mental health issues they target. This information was mostly presented in the titles, although sometimes it was more implicit, e.g., in the data collection and intervention techniques used. Table 1 presents the results and the answer to RQ1.

**Table 1.** Answering RQ1. Which mental health issues do the systems target?

| Work  | Targeted Mental Health Issue                                   |
|-------|--|
| [37]  | Depression   |
| [38]  | General well-being   |
| [114] | General well-being, stress, anxiety, depression                |
| [115] | General well-being, stress, anxiety, depression, loss of focus |
| [43]  | General well-being, stress                                     |
| [116] | Stress (in students)   |
| [117] | Depression   |
| [118] | General well-being   |
| [120] | (Occupational) stress  |
| [84]  | See [120]  |

#### 3.4.2. RQ2. Which Technologies, Methods and Collected Data Guide the Process to Achieve ABC for SAD in the Systems?

To answer RQ2, we scanned the reviewed works for information on which data was collected from the users by the authors and their systems, which datasets the authors used to train or augment their systems, what methods the systems were built on to produce ABC for SAD, and overall technologies used. All the listed had to have a specific purpose in producing ABC as opposed to, e.g., general conversational abilities of the system. There is a general process in treating mental health issues, which widely consists of two steps: assessment and intervention [121,122]. This was our further framework through which we viewed the reviewed works when looking for the answer to RQ2. Therefore, Table 2 presents the results and a part of the answer to RQ2 in regards to the assessment capabilities



of the reviewed systems, while Table 3 presents the results and a part of the answer to RQ2 in regards to the intervention capabilities of the reviewed systems.

**Table 2.** Answering RQ2. Which technologies, methods and collected data guide the process to achieve ABC for SAD in the systems? First step in the process: Assessment.

| Work  | Assessment  |
|-------|---|
| [37]  | The system tries to classify depression, suicidal ideation, insomnia and hypersomnia, weight change, and excessive or inappropriate guilt from linguistic user input. It trains on various datasets (eRisk, Reddit posts from users and subreddits). It extracts linguistic features from the text and uses doc2vec to vectorize it, employing feature recognition and text embedding approach to construct classifiers. It finally applies Random Forest and logistic regression to predict the presence or absence of depression symptoms. The overall F1-Score for classifiers was 0.91.   |
| [38]  | The system uses a lexicon-based approach using SentiWS lexicon to conduct sentiment and emotion recognition in the linguistic user input. It further applies fuzzy matching to recognize emotions from words that are similar enough to convey the same meaning. The system achieved 81% accuracy in recognizing emotions in a dataset of forum posts.  |
| [114] | The system collects geolocation data from a phone, connected to the ICA, and user ID, gender, baseline scores of the big five personality test, PANAS (Positive and Negative Affect Scale, short version), and DASS (Depression, Anxiety and Stress Scale). PANAS quantifies mood and DASS captures depression, anxiety, and stress symptoms. It applies experience sampling five times a day using a visual grid based on Russel's two-dimensional model of emotion to capture ground-truth labels. The affect is inferred by the system using personalized model with Random Forest regression for valence prediction (82.4% accuracy) and AdaBoost regression for arousal prediction (65.7% accuracy). |
| [115] | The system does not explicitly assess users and uses no specific assessment methods. Assessment is implicit in the linguistic intent recognition in the conversational model.   |
| [43]  | The system does not explicitly assess users and uses no specific assessment methods. Assessment is implicit in its matching capabilities where the user input is matched with the closest reply from the used database.   |
| [116] | The system does not explicitly assess users and uses no specific assessment methods. It uses evocative questions to collect linguistic user input. Afterwards, it uses keywords from the linguistic user input which guide the conversation—these keywords can convey mental states. The keywords were acquired from a dataset that collected data from Reddit subreddits.  |
| [117] | The system uses questions that target emotional states of the users to gain relevant user input. It then uses a model to detect seven types of emotions. The model uses long-short-term-memory neural network with glove2 for emotion recognition. The model is trained on the ISEAR dataset. The accuracy of emotion recognition obtained was 84%. Furthermore, it labels users into five states according to the detected emotional levels: zero depression, slightly stressed, highly stressed, slightly depressed, highly depressed.  |
| [118] | The systems do not explicitly assess users and use no specific assessment methods. Assessment is either implicit in the linguistic intent recognition in the conversational model or it uses keywords from the linguistic user input which guide the conversation.  |
| [120] | The system uses fuzzy inference to evaluate the content of the linguistic user input as replies to various intentional questions and to detect users' state of stress. The users are measured on Comprehensibility (Co), Manageability (Ma) and Meaningfulness (Me). "Comprehensibility means that people can understand their situation and predict their near future. Manageability is a sense that people can manage their situation. Meaningfulness means people can understand the meaning of their life." [120] (p. 3763). This determines users' Sense of Coherence (SOC) model, which is used for various strategies to increase stress management.   |
| [84]  | See [120]   |

**Table 3.** Answering RQ2. Which technologies, methods and collected data guide the process to achieve ABC for SAD in the systems? Second step in the process: Intervention.

| Work  | Intervention   |
|-------|--|
| [37]  | The system does not deliver interventions and has not specific intervention methods.   |
| [38]  | The system delivers suggestions for activities and exercises that help regulate emotions in the form of a dialogue, reminds the user on appointments and implements CBT techniques, e.g., mindfulness and focusing in goals. The dialogues vary depending on detected emotions and are mostly of informational nature.   |
| [114] | The system delivers well-being interventions which include individual or social activities from a range of psychotherapeutic categories: positive psychology, cognitive behavioral, meta-cognitive, or somatic interventions. They are delivered through a textual prompt to the user with various digital tools to engage with the activity. The dialogue the system produces is based on emotions detected by selecting a random pre-written script from an emotional category, congruent with the user's state (e.g., if a person is identified to have emotions of low valence and arousal, the system produces the following: "Feeling glum? I have a skill that might brighten your day. Let us practice.").   |
| [115] | The system delivers interventions in the form of positive drivers inserted in the conversation to change the trend of the users' thoughts. It also targets self-expression development and stress management. CBT techniques, motivational interviewing and analysis, positive behavior support, behavioral reinforcement, and guided actions and methods are used to encourage the user to build emotional resilience skills. Actions are encouraged at different moments, such as meditation.  |
| [43]  | The system delivers interventions in the form of preexisting emotional support statements, drawn from a large corpus of online interactions from the Koko platform, a platform that connects users seeking help and those who have opted to give help. The users needing help also evaluate the responses. This corpus-based approach tries to create the semblance of personalized, empathic expression. The system uses information retrieval techniques and word embeddings to automate this process in real-time, matching existing statements to appropriate inputs by the users, selecting texts that have satisfactory scores. The interaction between the system and the user is one-off—the user describes her situation and the system matches a reply from the dataset. The answers are presented as if authored by the system. |
| [116] | The system delivers interventions in the form of motivational interviewing. It can only use predefined responses, which depend on the stage of the process the user is in. These stages are Engaging, Focusing, Evoking, and Planning, where: "In Engaging, Bonobot shares brief introductions with the user and gives instructions to use the chatbot. In Focusing, Bonobot asks the user to detail their problem, possibly having them identify an inner struggle. This leads to Evoking, where Bonobot explores future goals with the user, affirming their own ideas for change. Finally, Bonobot invites the user to ponder the overall session in Planning." [116] (p. 3). The process helps users cope with stress and encourages self-reflection.  |
| [117] | The system delivers interventions in the form of emotional conversational support, suggesting different, more positive perspectives on situations the users describe, and trying to prevent negative thoughts. The conversation is guided by the level of mental health issue detected.  |
| [118] | The systems deliver interventions differently. ELIZA-based ICA uses Rogerian reflection to engage with the users. Information retrieval techniques are used to choose proper responses: the n-gram technique, charagram embeddings, word similarity, sentence similarity, and part-of-speech tagging. ALICE-based ICA delivers interventions by sympathizing with the user and using CBT techniques. It implements AIML, sklearn to match responses, as well as category tagging and synonym switching for conversational dynamicity. The generative language ICA only implicitly delivers interventions by being trained on empathetic text.  |

Table 3. Cont.

| Work  | Intervention   |
|-------|--|
| [120] | The system delivers interventions that help improve the users' self-efficacy, which helps manage stress, as it measures users' sense of task performance and whether they feel they can do a task or not. The system, drawing from the user model which is based on the SOC model, engages the Peer Support model, which finds suitable support types and delivers them. The system uses reinforcement learning and fuzzy control to find the best Peer Support types for specific SOC models. Peer support also stimulates various aspects of a person to lower stress levels. The types of support are helper therapy (the user takes the role of the carer instead of being cared for), informational support, esteem support, and emotional support. |
| [84]  | See [120]. The authors upgraded the system with expanding the helper therapy support type by the user having to be a carer offering either informational or emotional support, depending on their SOC.   |

### 3.4.3. RQ3. What Are the Technical Aspects of the Conversational Models in the Systems?

To answer RQ3, we scanned the reviewed works for information on which methods were used to build conversational models in the reviewed systems. Generally, we see two approaches: rule-based, dialogue tree conversational models with either free text or button-based user input options (more control, less errors, but limited conversational experiences), and generative models with free text options (less control, more errors, more affordances for conversation). Table 4 presents the results and the answer to RQ3.

Table 4. Answering RQ3. What are the technical aspects of the conversational models in the systems?

| Work  | Conversational Model   |
|-------|--|
| [37]  | The system's conversational mode was trained with seq2seq (OpenNMT) learning approach on datasets from Reddit's subreddits, the eRisk dataset and OpenSubtitles dataset using neural networks.   |
| [38]  | The system's conversational model is built on the Syn.Bot framework, which uses Oscova as the bot development platform and the SIML (Synthetic Intelligence Markup Language) interpreter. The model lets the users frame answers in their own words and select predefined answers.   |
| [114] | The system's conversational model works on textual prompts and scripted phrasings that are utilized at contextually appropriate times.   |
| [115] | The system's conversational model uses RNNs for learning as well as understanding and generating responses. The intent in the user input is recognized by the Long-Short-Term-Memory neural network.   |
| [43]  | The system's conversational model consists of two modules. The front-end module pairs previous responses with user inputs. The back-end module generates output using Elasticsearch, word2vec and a word-embedding procedure. The authors used the Google News dataset for training. The ICA also solicits user feedback.  |
| [116] | The system's conversational model extends on ELIZA, basing its functionalities on identifying user keywords to generate responses. It consists of two modules, Flow Manager and Response Generator. Flow Manager runs the conversation and assigns template responses to lead the user. Response Generator follows the conversational flow and sequences, identifying keywords by weighting them and assembling responses. |
| [117] | The system's conversational model is built by the authors using word embeddings, word2vec, glove, pre-written questions, and trained responses to create an environment for generative, free-text conversation.  |

Table 4. Cont.

| Work  | Conversational Model   |
|-------|--|
| [118] | The three systems' conversational models are built with three different approaches: (1) the Retrieval Pattern Matching ICA is built on ELIZA, using the n-gram technique to get relevant responses, Charagram embeddings to learn character-based compositional models to embed textual sequences, and using word similarity, sentence similarity and part-of-speech tagging for evaluation; (2) Retrieval Rule Based AIML ICA is built on ALICE, using sklearn alongside the AIML library and various rules to generate a response; (3) the generative ICA learns on the data from The Open American National Corpus, using the Long-Short-Term-Memory method and context learning for understanding input, and using Beam Search to choose a response. |
| [120] | The system's conversational model is rule-based, basing its responses on a stored databank. The user can communicate by inputting free text or by selecting fixed inputs. The outputs are also based on the classification of the moods of the users, detected through using machine learning (see Table 2).   |
| [84]  | See [120]  |

#### 3.4.4. RQ4. What Are the Platforms Used to Create the Systems?

To answer RQ4, we scanned the reviewed works on how the ICAs were built. We focused on whether various platforms were used to produce the ICA (e.g., Rasa [123]) or whether an existing ICA and its framework were used and possibly upgraded (e.g., ELIZA). As this was one of our exclusion criteria, we did not consider papers with ICAs built on proprietary, closed code platforms (e.g., DialogFlow). Table 5 presents the results and the answer to RQ4.

Table 5. Answering RQ4. What are the platforms used to create the systems?

| Work  | Platforms and Frameworks   |
|-------|--|
| [37]  | No existing platform or framework/Not reported   |
| [38]  | Syn.Bot, OSCOVA  |
| [114] | StudyPortal platform (extricated from [124])   |
| [115] | No existing platform or framework/Not reported   |
| [43]  | No existing platform or framework/Not reported   |
| [116] | Extended ELIZA framework   |
| [117] | No existing platform or framework/Not reported   |
| [118] | Extended ELIZA framework, extended ALICE framework; no platform/framework reported for the third ICA |
| [120] | No existing platform or framework/Not reported   |
| [84]  | LINE Platform  |

#### 3.4.5. RQ5. What Domain Knowledge Is Used to Achieve ABC for SAD?

To answer RQ5, we scanned the reviewed works on what domain knowledge, particularly from mental health and ABC theories, is somehow integrated into the systems. This may be through the strategies that the systems deploy to produce ABC, e.g., CBT techniques, or through user modeling, where knowledge on SAD helps make the systems more empathetic. Table 6 presents the results and the answer to RQ5.

**Table 6.** Answering RQ5. What domain knowledge is used to achieve ABC for SAD?

| Work  | Domain Knowledge  |
|-------|---|
| [37]  | See Table 2   |
| [38]  | The system reflects knowledge on self-reflection, tracking, monitoring (diaries), ABC theory, information provision, and CBT techniques like mindfulness and goal-attainment.   |
| [114] | The system reflects knowledge on positive psychology, cognitive behavioral, meta-cognitive, or somatic interventions as well as emotion theory like Russel's circumplex model.  |
| [115] | The system reflects knowledge on “self-help practices such as CBT, motivational interviewing and analysis, positive behavior support, behavioral reinforcement and guided actions and methods to encourage the user to build emotional resilience skills. It helps the user to manage their stress, anxiety, overthinking, energy, helps in focus, promotes meditation and encourages the same, and other situations.” [115] (p. 122) |
| [43]  | The system reflects no explicitly discernible domain knowledge.   |
| [116] | The system reflects knowledge on motivational interviewing, stress management, and self-reflection.   |
| [117] | The system reflects knowledge on the emotion theory (seven basic emotions), emotional support and evocative questions.  |
| [118] | The system reflects knowledge on Rogerian reflection.   |
| [120] | The system reflects knowledge on the SOC model, Generalized Resistance Resources, helper therapy, informational support, and emotional support.   |
| [84]  | See [120]   |

#### 3.4.6. RQ6. What User Modeling, Especially for Personalization and Adaptation, Do the Systems Conduct?

To answer RQ6, we scanned the reviewed works on what kind of data is collected on the users for the user model, and how the user is further modeled. We were also interested in how this affects the working of the system, especially in terms of how the system is personalized and how it adapts to individual users. Table 7 presents the results and the answer to RQ6.

**Table 7.** Answering RQ6. What user modeling, especially for personalization and adaptation, do the systems conduct?

| Work  | User Modeling   |
|-------|---|
| [37]  | See Table 2   |
| [38]  | The system builds the user model on the emotion data, which it uses to personalize dialogues.   |
| [114] | The system builds the user model on the following data: “user ID, gender, baseline scores of the big five personality test, PANAS (Positive and Negative Affect Scale, short version), and DASS (Depression, Anxiety and Stress Scale). PANAS quantifies mood and DASS captures depression, anxiety, and stress symptoms.” [114] (p. 16). It also contains data on “experience sampling five times a day using a visual grid based on Russel’s two-dimensional model of emotion.” [114] (p. 16). It uses this data to select among different emotionally charged phrasings. |
| [115] | The system does not build any explicit user models.   |



**Table 7.** *Cont.*

| Work  | User Modeling   |
|-------|---|
| [43]  | The system does not build any explicit user models.   |
| [116] | The system does not build any explicit user models.   |
| [117] | See Table 2   |
| [118] | The systems do not build any explicit user models.  |
| [120] | The system builds the user model on the following data: data from the SOC model, Perceived Stress Scale, Ryff's Psychological Well-Being Scales, and Hassles Scale. Each user has continually updated SOC model. Generalized Resistance Resources connect other data to the SOC mode. |
| [84]  | See [120]   |

### 3.4.7. RQ7. What Is the Overarching Cognitive Architecture Used in the Systems?

To answer RQ7, we scanned the reviewed works to see if they refer to any kind of specific, pre-defined cognitive architecture (e.g., Belief-Desire-Intention architecture) they followed when constructing the system. If they did not, we were interested to see which modules comprise the cognitive architecture. Table 8 presents the results and the answer to RQ7.

**Table 8.** Answering RQ7. What is the overarching cognitive architecture used in the systems?

| Work  | Cognitive Architecture  |
|-------|---|
| [37]  | Not specified; modules for various mental issue problems detection, conversational model for question formation   |
| [38]  | Syn-Bot architecture (including OSCOVA and SIML)  |
| [114] | Not specified; geolocation-emotion prediction module, personalized textual interventions module   |
| [115] | Not specified; language learning module (RNN), user understanding module (NLP), response generator (NLP) with psychological techniques  |
| [43]  | Not specified; pairing module, user feedback module   |
| [116] | Not specified; flow manager, response generator   |
| [117] | Not specified; mental state classification module, response generator, user model   |
| [118] | Not specified; ELIZA-based system: pattern matching module, response generator; ALICE-based system: self learning module, response generator; Generative system: training module, context module, generalization module, response generator |
| [120] | Belief-Desire-Intention architecture  |
| [84]  | See [120]   |

### 3.4.8. RQ8. How Are the Systems Evaluated in Terms of ABC for SAD?

To answer RQ8, we scanned the reviewed works to see the evaluation of the systems, focusing on user tested evaluation. Ideally, we wanted to see the mental health outcomes after using the system, but we also extracted data on user evaluation in terms of evaluating the system's properties. Table 9 presents the results and the answer to RQ8.

**Table 9.** Answering RQ8. How are the systems evaluated in terms of ABC for SAD?

| Work  | Evaluation   |
|-------|--|
| [37]  | No evaluation on users   |
| [38]  | Tested on users and mental health professionals on the system's Attractiveness (users: below average; professionals: good), Perspicuity (users: above average; professionals: above average), Efficiency (users: below average; professionals: above average), Dependability (users: bad; professionals: below average), Stimulation (users: bad; professionals: above average), and Novelty (users: below average; professionals: excellent). |
| [114] | No evaluation on users   |
| [115] | No evaluation on users   |
| [43]  | Tested on users where they compared the system's replies to their peers' replies with three scores: <i>good</i> (system: >40%; peers: >60%), <i>ok</i> (system: <40%; peers: <40%), and <i>bad</i> (system: >20%; peers: <10%).  |
| [116] | Tested on users where they described the system as having evocative questions and offering self-reflection as well as potential consolidation, but noted that the feedback was clichéd. The users also wanted more informational support from the system and more suitably contextualized feedback.  |
| [117] | No evaluation on users   |
| [118] | No evaluation on users   |
| [120] | Tested on users which used the system for five days. The system managed to improve their scores on stress managing skills, reflected in the SOC model, and their stress levels fell (approx. 30% improvement).   |
| [84]  | Tested on users which used the system for 3 days. The system managed to improve their scores on stress managing skills, reflected in the SOC model.  |

The answers to the research questions give a thorough and detailed insight into how the reviewed systems produce ABC for SAD, especially in their underlying technical mechanisms. This is especially relevant to see what kind of data should be collected on users, how they should be modeled to personalize and adapt ICAs, how the latter should converse with the users, etc. The tables with results, which allow for easy comparisons, tell a story into what the current SOTA on how to produce change in stress, anxiety and depression with autonomous dialogue systems is. The following section discusses our work in comparison to the reviews in Section 1.5, and our results from answering the RQs, especially in the light of their significance in the wider technological landscape.

#### 4. Discussion

What we have discerned with this comprehensive review is that the technically inclined research community for the reviewed systems is not large. Consequently, the discussions that can be had at this point are necessarily limited. Nevertheless, this section tries to weave a wider narrative on ICAs for ABC in mental health, backed by the results from the previous section.

##### 4.1. Comparison of Existing Reviews

This review fundamentally differs from existing reviews (covered in Section 1.5) to the degree where we are confident calling it the first review of its kind. While the other reviews cover similar technologies, their research questions and selection criteria were entirely different. Generally, they focused on delivering a systematized review for health practitioners, which molded their research questions in the direction of looking at outcomes of using such system in terms of how they influence the users' mental health. They were evaluating the possible benefits such system could have if used in mental healthcare. Due

to such focus, their selection criteria was not interested in whether there is enough technical information on the system overviewed to analyze it. This meant that they included mostly proprietary, commercial systems, which give no insight into how they are built. Such reviews, although immensely valuable for the interdisciplinary research area, does not provide the knowledge that people included in the system development could use to further advance the current technological landscape. Our work therefore stands alone in reviewing and systematizing the trends of currently non-proprietary ICAs for ABC and SAD symptom relief. We believe this work will be helpful for researchers developing such systems to base their efforts on, get ideas, and potentially find communities. The paper may also serve health professionals to get acquainted with the technology they might be using in the future, and to better understand it, potentially increasing their trust in introducing technology into mental healthcare.

#### *4.2. Comparison of Systems from Selected Papers*

The approaches to ABC, observed in the reviewed systems, considerably vary. It benefits to compare the technical underpinnings of systems targeting the same mental health issue.

Targeting stress, both systems by Yorita et al. [84,120] produced experimental results in reducing stress symptoms comparable to SOTA results, observed in review papers in Section 1.5. We believe the system achieved this by: having strong theoretical grounds for assessment, which produced comprehensive user models of the users' stress management skills as well as other psychological aspects; having explicitly personalized interventions, which were selected from a wide range of possibilities according to the factors in the user model; having a rule-based conversational model, which guided the user down appropriate dialogue paths instead of having the freedom to go off-topic (or down erroneous paths) as in free text conversational models; basing its domain knowledge on a few carefully selected psychological frameworks, such as SOC model, helper therapy, informational support, and others; and choosing a well-supported cognitive architecture, Belief-Desire-Intention architecture, to build the system on. Other systems targeting stress lacked such comprehensive architecture in terms of its modules. Some built comprehensive user models but lacked the depth of personalized strategies rooted in theory, opting for few pre-written responses [114]; some did not explicitly assess and intervene, opting for approaches that are more dependent on unsupervised understanding of and responding to users [115,118]; some produced very rigid and static systems based on a lot of top-down elements to assessment and intervention, either through matching with already existing responses [116] or by following a very strict and limited conversational path [43]. It therefore seems that a strong user model with an intelligent combination of rigidness and freedom of assessment and intervention methods through a guided conversation produces best results.

Targeting anxiety, no systems with experimental results targeting symptom reduction were found. Two systems targeted anxiety, but very generally, either through few pre-written responses [114], or by opting for dialogue freedom through a generative conversational model [115]. Ghandeharioun et al. [114], however, built their system technically based on assessment, using Random Forest and AdaBoost with satisfying results to infer mood from a comprehensive user data model, which might be a better option than implicit assessment.

Targeting depression, Delahunty et al. [37] presented the most comprehensive system for depression assessment building various classifiers on depression symptoms used on the input text. Random Forest and logistic regression were used to infer the presence of depression, suicidal ideation, insomnia and hypersomnia, weight change, and excessive or inappropriate guilt. This appears to be a more nuanced way to assess users than opting for general mental health issue labels. However, their system was assessment only. Ghandeharioun et al. [114] and Khadikar et al. [115] systems were already covered in the previous paragraphs, and the same evaluation applies here.

Systems targeting general well-being are harder to compare, but Denecke et al. [38] seemed to follow the formula of Yorita et al. in terms of building a comprehensive system with the right combination of rigidity and dynamicity in assessment, intervention and guided conversation. The system's performance seemed to be based on their assessment methods, which used a lexicon approach to extract linguistic features and infer emotions in the text.

In summary, successful systems seem to base their performance on a comprehensive user model, explicit and theoretically-backed assessment with classification models (instead of only collecting questionnaire results), explicit and personalized intervention with many strategic possibilities, and dialogue tree conversational model. As in many areas, tasks that call for machine learning are best solved with ensemble methods, such as Random Forest.

#### 4.3. Technology Evaluation

Overall technological evaluation of the existing systems is harder due to the usage prevalence of proprietary systems. ICAs like Woebot [17], Tess [18] and Wysa [19] seem to possess architectures with SOTA ABC capabilities for SAD, and it is a shame that we were not able to include them in our research.

There are a few clear insights into the preferable technologies that the reviewed ICAs are built on. The first noticeable element is the intricate connection between the technology and the goals of such ICAs. Here, it can be discerned that conversational models in most cases are built to be fairly limited in what is otherwise SOTA in the field of chatbots. It has to be limited—mental health counselling is a very delicate matter, and preventing the generative models go out of control should be one of the primary concerns, as making them be complicit in mental health deterioration of the user is a real danger. This was seen in the case of the currently most advanced language model today, OpenAI's GPT-3 [112]. GPT-3 was being tested by the tester simulating a patient. When the tester simply wanted to book an appointment with a doctor, GPT-3 acted as a human, understanding the tester's intents with no problems. However, beyond such surface tasks and conversations, GPT-3 started not only to fault, but to exhibit very dangerous behaviors. When the tester expressed that she feels bad and needs help, GPT-3 answered that it can help, and when the tester expressed suicidal thoughts, GPT-3 recommended that the tester killed herself [125]. That this occurred with the most advanced language model in the world, produced by the leading AI research organization, is worrying to say the least. To researchers in this field, it signals not only how careful they have to be, but also that the systems they build have to be very domain-oriented and should limit the linguistic capabilities as reasonably as possible. In the domain of mental health, it is clear that free text capabilities of ICAs are not on the level where they could be feasibly used, and that generally, NLP research is not advanced enough yet to consider it for such domains [126]. When they are used, they have to be largely improved on in very domain-specific ways, making the systems non-scalable.

While the authors of the reviewed works were aware of the dangers of unconstrained textual input, their conversational models seemed too limited in what is currently possible. One glaring omission that the current language technologies feasibly offer, at least to explore and make progress on, is that the conversational model can remember historical interactions with the same user. This enables a more long-term connection between the ICA and the user, where the therapy has so much more possibilities to explore. The bond that forms and the information that can be gathered can produce much better outcomes. One possible reason why the authors did not implement this is convenience and privacy—the user does not have to create an account, which removes some initial barriers to the system use, and the system does not have to store any historical data on the user, which enhances privacy.

The latter may also be the reason why there is so little user modeling and consequential personalization. The systems collect very little data on the users, which makes them static and inflexible in terms of how they can personalize their strategies to the user and adapt to various individual specificities. Since the current systems do seem to employ ABC

theories and strategies, personalizing offered help to specific groups that are affected more by specific strategies [127] should be the logical next step in progressing these systems.

Due to the conversational models many times being the most fleshed out part of the reviewed ICAs, their cognitive architectures are not thought out in high detail, sometimes embedding only the conversational model. This can cause oversimplification of possibilities for the system to function, which has its place for certain purposes (very general and quick first help), but does not explore the possibilities that modeling other cognitive capabilities can bring. Over-reliance on conversational models has another downside: most work well (where they work well) for the English language [128], but hardly for other languages. Lexicons for relevant feature extraction and language datasets for training in non-English language are few compared to how many exist in English. Opting for anything other than English hinders the possibility to produce SOTA capabilities of the explored systems. Another downside of this is that non-English speakers cannot use the majority of systems produced.

Some designs of ICA cognitive architectures [31] have suggested how to sensibly use more advanced technology which might result in better outcomes, but have so far not been implemented or evaluated yet. They emphasize personalization and adaptation through strong user modeling and learning from historical interactions. It is clear that ICAs for ABC in mental health have a lot of space to grow technologically, should there be enough research in the field. The most important lesson to note is that the outcomes such ICAs produce are emergent—they represent a thoroughly researched and thought out result of highly interdisciplinary efforts, but more specifically, their behavior stems from various modules that model different cognitive abilities interacting with each other. This points to researchers needing to cooperate or being interdisciplinary themselves, not only focusing on narrow intradisciplinary or technical knowledge.

## 5. Conclusions and Future Work

This state-of-the-art technical review presents the first technical review of intelligent cognitive assistants that produce attitude and behavior change for people with stress, anxiety and depression symptoms. It introduces the topic of change and its importance as the holy grail of different research fields and human endeavors, lays out our motivation for the work, and continues to describe the interdisciplinary connections between attitude and behavior change support systems, intelligent cognitive assistant technology, and digital mental health. It presents related works—similar reviews, but points to these not being technical and targeting health practitioners, which can be discerned from the lack of technological analysis of the systems. The work further lays out our methodology, presents the process of finding and selecting papers and, finally, presents the results, which are put into context in the discussion. The results tell a story of how various systems try to achieve change, employing various technological and scientific mechanisms. However, these systems do not reflect the possible SOTA, which can be achieved with more research.

The biggest limitation of this work, as already addressed, concerns the lack of inclusion of various proprietary systems, which would bring additional value to the technical analysis this paper offers. Another limitation might be the specific criteria we constructed for the paper selection. Although we tried to produce the criteria non-arbitrarily, providing reasons for our decisions, some important papers to present might not have been included in this work. We must also consider that papers that would fit the criteria might be only be available in some smaller, specific databases that we did not include in our search. What was also limiting was our focus on the most common mental health issues, and mental health issues that such technologies usually target, especially since they are mostly experienced among non-clinical population. Including other mental health issues would widen the scope, meaning that papers with systems targeting these mental health issues could include technologies not covered in this work. The final identified limitation concerns covered related work. We focused on reviews that focused on outcomes of using reviewed systems, but it would be worthwhile to explore reviews of such systems that focus on some



other aspect, e.g., acceptability, convenience of use, adherence, and data protection and privacy solutions.

Our future work is guided by the limitations listed. Including different kinds of systems for attitude and behavior change in mental health is needed to explore how technologies, not covered here, might prove beneficial. Including other mental health issues (i.e., autism, psychosis) is needed to explore how certain technologies might only work for certain mental health issues. Surveying the suggestions for systems' designs is also something to consider as to consolidate various lessons learned.

The novel contribution that this review represents points to the still emerging research field that is gaining prominence due to the ubiquity of technology and the rise of mental health issues. With meaningful integration with the existing mental healthcare and further research, artificial systems might play an important role in bettering the current mental landscape.

**Author Contributions:** Conceptualization, T.K.; methodology, T.K.; software, T.K.; validation, T.K. and M.G.; formal analysis, T.K. and M.G.; investigation, T.K. and M.G.; resources, M.G.; data curation, T.K.; writing—original draft preparation, T.K.; writing—review and editing, T.K. and M.G.; visualization, T.K.; supervision, M.G.; project administration, T.K.; funding acquisition, M.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Slovenian Research Agency (research core funding No. P2-0209 and Young researchers postgraduate research funding).

**Acknowledgments:** The authors acknowledge the financial support from the Slovenian Research Agency (research core funding No. P2-0209 and Young researchers postgraduate research funding).

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

## Abbreviations

The following abbreviations are used in this manuscript:

|       |   |
|-------|---|
| ABC   | Attitude Furthermore, Behavior Change   |
| AI    | Artificial Intelligence                 |
| AIML  | Artificial Intelligence Markup Language |
| CBT   | Cognitive Behavioral Therapy            |
| CPP   | Cialdini's Principles of Persuasion     |
| FBM   | Fogg Behavior Model                     |
| DASS  | Depression, Anxiety and Stress Scale    |
| ECA   | Embodied Conversational Agent           |
| ICA   | Intelligent Cognitive Assistant         |
| PANAS | Positive and Negative Affect Scale      |
| PSDM  | Persuasive System Design Model          |
| PT    | Persuasive Technology                   |
| RNN   | Recurrent Neural Network                |
| SAD   | Stress, Anxiety Furthermore, Depression |
| SIML  | Synthetic Intelligence Markup Language  |
| SOC   | Sense Of Coherence                      |

## References

1. Seibt, J. Process Philosophy. In *The Stanford Encyclopedia of Philosophy*, 2020 ed.; Zalta, E.N., Ed.; Metaphysics Research Lab, Stanford University: Stanford, CA, USA, 2020.
2. Wikipedia Contributors. Confucianism—Wikipedia, The Free Encyclopedia. 2021. Available online: <https://en.wikipedia.org/wiki/Confucianism> (accessed on 20 April 2021).
3. Dziurosz-Serafinowicz, D. Aquinas' concept of change and its consequences for corporeal creatures. *Logos I Ethos* **2014**, *1*, 173. [CrossRef]

4. Watzlawick, P.; Weakland, J.; Fisch, R. *Change: Principles of Problem Formation and Problem Resolution*; Norton: New York, NY, USA, 1974.
5. Wikipedia Contributors. Empedocles—Wikipedia, The Free Encyclopedia. 2021. Available online: <https://en.wikipedia.org/wiki/Empedocles> (accessed on 20 April 2021).
6. Martin, S.; Goldstein, N.; Cialdini, R. *The Small BIG: Small Changes That Spark Big Influence*; Grand Central Publishing: New York, NY, USA, 2014.
7. van Veen, V.; Krug, M.K.; Schooler, J.W.; Carter, C.S. Neural activity predicts attitude change in cognitive dissonance. *Nat. Neurosci.* **2009**, *12*, 1469–1474. [[CrossRef](#)] [[PubMed](#)]
8. Gams, M.; Kolenik, T. Relations between Electronics, Artificial Intelligence and Information Society through Information Society Rules. *Electronics* **2021**, *10*, 514. [[CrossRef](#)]
9. Watanabe, M.E. The United Nations Sustainable Development Goals: Researchers seek to make headway amid obstacles. *BioScience* **2020**, *70*, 205–212. [[CrossRef](#)]
10. Auerbach, J.; Miller, B.F. COVID-19 Exposes the Cracks in Our Already Fragile Mental Health System. *Am. J. Public Health* **2020**, *110*, 969–970. [[CrossRef](#)]
11. Provoost, S.; Lau, H.M.; Ruwaard, J.; Riper, H. Embodied Conversational Agents in Clinical Psychology: A Scoping Review. *J. Med. Internet Res.* **2017**, *19*, e151. [[CrossRef](#)]
12. Vaidyam, A.N.; Wisniewski, H.; Halamka, J.D.; Kashavan, M.S.; Torous, J.B. Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape. *Can. J. Psychiatry* **2019**, *64*, 456–464. [[CrossRef](#)]
13. Abd-Alrazaq, A.A.; Rababeh, A.; Alajlani, M.; Bewick, B.M.; Househ, M. Effectiveness and Safety of Using Chatbots to Improve Mental Health: Systematic Review and Meta-Analysis. *J. Med. Internet Res.* **2020**, *22*, e16021. [[CrossRef](#)]
14. Gaffney, H.; Mansell, W.; Tai, S. Conversational Agents in the Treatment of Mental Health Problems: Mixed-Method Systematic Review. *JMIR Ment. Health* **2019**, *6*, e14166. [[CrossRef](#)]
15. Bendig, E.; Erb, B.; Schulze-Thuesing, L.; Baumeister, H. The Next Generation: Chatbots in Clinical Psychology and Psychotherapy to Foster Mental Health—A Scoping Review. *Verhaltenstherapie* **2019**. [[CrossRef](#)]
16. Abd-alrazaq, A.A.; Alajlani, M.; Alalwan, A.A.; Bewick, B.M.; Gardner, P.; Househ, M. An overview of the features of chatbots in mental health: A scoping review. *Int. J. Med. Inform.* **2019**, *132*, 103978. [[CrossRef](#)]
17. Fitzpatrick, K.K.; Darcy, A.; Vierhile, M. Delivering Cognitive Behavior Therapy to Young Adults with Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. *JMIR Ment. Health* **2017**, *4*, e19. [[CrossRef](#)]
18. Fulmer, R.; Joerin, A.; Gentile, B.; Lakerink, L.; Rauws, M. Using Psychological Artificial Intelligence (Tess) to Relieve Symptoms of Depression and Anxiety: Randomized Controlled Trial. *JMIR Ment. Health* **2018**, *5*, e64. [[CrossRef](#)] [[PubMed](#)]
19. Inkster, B.; Sarda, S.; Subramanian, V. An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study. *JMIR Mhealth Uhealth* **2018**, *6*, e12106. [[CrossRef](#)] [[PubMed](#)]
20. Fogg, B. *Persuasive Technology: Using Computers to Change What We Think and Do*; Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA, 2003.
21. Orji, R.; Moffatt, K. Persuasive technology for health and wellness: State-of-the-art and emerging trends. *Health Inform. J.* **2018**, *24*, 66–91. [[CrossRef](#)] [[PubMed](#)]
22. Cialdini, R.B. *Influence: Science and Practice*; Pearson Education: Boston, MA, USA, 2009.
23. Kahneman, D. *Thinking, Fast and Slow*; Farrar, Straus and Giroux: New York, NY, USA, 2011.
24. Thaler, R.; Sunstein, C. *Nudge: Improving Decisions about Health, Wealth, and Happiness*; Yale University Press: New Haven, CT, USA, 2008.
25. Berghel, H. Malice Domestic: The Cambridge Analytica Dystopia. *Computer* **2018**, *51*, 84–89. [[CrossRef](#)]
26. United Nations Sustainable Development—17 Goals to Transform Our World. Available online: <https://www.un.org/sustainabledevelopment/> (accessed on 30 September 2020).
27. Midden, C.; Mcalloy, T.; Ham, J.; Zaalberg, R. Using persuasive technology to encourage sustainable behavior. *Sustain. WS Pervasive* **2008**, *113*, 83–86.
28. Gram-Hansen, S.; Svarre, T.; Midden, C. Persuasive Technology. Designing for Future Change. In Proceedings of the 15th International Conference on Persuasive Technology (PERSUASIVE 2020), Aalborg, Denmark, 20–23 April 2020. [[CrossRef](#)]
29. Oinas-Kukkonen, H.; Harjumaa, M. Persuasive Systems Design: Key Issues, Process Model, and System Features. *Commun. Assoc. Inf. Syst.* **2009**, *24*. [[CrossRef](#)]
30. Gkika, S.; Skiada, M.; Lekakos, G.; Kourouthanassis, P.E. Investigating the Role of Personality Traits and Influence Strategies on the Persuasive Effect of Personalized Recommendations. *EMPIRE@RecSys*. 2016. Available online: <http://ceur-ws.org/Vol-1680/paper2.pdf> (accessed on 23 April 2021).
31. Kolenik, T.; Gams, M. PerMEASS—Personal Mental Health Virtual Assistant with Novel Ambient Intelligence Integration. *CEUR-WS*. 2020. pp. 8–12. Available online: <http://ceur-ws.org/Vol-2820/AAI4H-2.pdf> (accessed on 23 April 2021).
32. Michener, H.; DeLamater, J.; Myers, D. *Social Psychology*; Available Titles CengageNow, Wadsworth/Thomson Learning: Boston, MA, USA, 2003.
33. Rammstedt, B.; John, O.P. Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *J. Res. Personal.* **2007**, *41*, 203–212. [[CrossRef](#)]

34. Lee, K.; Ashton, M.C. HEXACO Model of Personality Structure. In *Encyclopedia of Personality and Individual Differences*; Zeigler-Hill, V., Shackelford, T.K., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 1932–1936. [\[CrossRef\]](#)
35. Lovibond, S.; Lovibond, P. *Manual for the Depression Anxiety Stress Scales*; Psychology Foundation Monograph, Psychology Foundation of Australia: Sydney, Australia, 1996.
36. Ratcliffe, M. *Experiences of Depression: A Study in Phenomenology*; International Perspectives in; Oxford University Press: Oxford, UK, 2015.
37. Delahunty, F.; Wood, I.D.; Arcan, M. First Insights on a Passive Major Depressive Disorder Prediction System with Incorporated Conversational Chatbot. In Proceedings of the 26th AIAI Irish Conference on Artificial Intelligence and Cognitive Science, Dublin, Ireland, 6–7 December 2018; pp. 327–338.
38. Denecke, K.; Vaaheesan, S.; Arulnathan, A. A Mental Health Chatbot for Regulating Emotions (SERMO)—Concept and Usability Test. *IEEE Trans. Emerg. Top. Comput.* **2020**. [\[CrossRef\]](#)
39. Gjoreski, M.; Kolenik, T.; Knez, T.; Luštrek, M.; Gams, M.; Gjoreski, H.; Pejović, V. Datasets for Cognitive Load Inference Using Wearable Sensors and Psychological Traits. *Appl. Sci.* **2020**, *10*, 3843. [\[CrossRef\]](#)
40. Laranjo, L.; Dunn, A.G.; Tong, H.L.; Kocaballi, A.B.; Chen, J.; Bashir, R.; Surian, D.; Gallego, B.; Magrabi, F.; Lau, A.Y.S.; et al. Conversational agents in healthcare: A systematic review. *J. Am. Med. Inf. Assoc.* **2018**, *25*, 1248–1258. [\[CrossRef\]](#)
41. Oakley, J. Intelligent Cognitive Assistants (ICA). 2018. Available online: [https://www.nsf.gov/crssprgm/nano/reports/ICA2\\_Workshop\\_Report\\_2018.pdf](https://www.nsf.gov/crssprgm/nano/reports/ICA2_Workshop_Report_2018.pdf) (accessed on 23 April 2021).
42. Wikipedia Contributors. ELIZA—Wikipedia, The Free Encyclopedia. 2021. Available online: <https://en.wikipedia.org/wiki/ELIZA> (accessed on 23 April 2021).
43. Morris, R.R.; Kouddous, K.; Kshirsagar, R.; Schueller, S.M. Towards an Artificially Empathic Conversational Agent for Mental Health Applications: System Design and User Perceptions. *J. Med. Internet Res.* **2018**, *20*, e10148. [\[CrossRef\]](#)
44. Cognitive Architecture. Available online: <http://cogarch.ict.usc.edu/> (accessed on 30 May 2020).
45. Garrod, S.; Pickering, M.J. Why is conversation so easy? *Trends Cogn. Sci.* **2004**, *8*, 8–11. [\[CrossRef\]](#)
46. Io, H.N.; Lee, C.B. Chatbots and conversational agents: A bibliometric analysis. In Proceedings of the 2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), Singapore, 10–13 October 2017; pp. 215–219. [\[CrossRef\]](#)
47. Lucas, G.M.; Gratch, J.; King, A.; Morency, L.P. It's only a computer: Virtual humans increase willingness to disclose. *Comput. Hum. Behav.* **2014**, *37*, 94–100. [\[CrossRef\]](#)
48. Salari, N.; Hosseini-Far, A.; Jalali, R.; Vaisi-Raygani, A.; Rasoulpoor, S.; Mohammadi, M.; Rasoulpoor, S.; Khaledi-Paveh, B. Prevalence of stress, anxiety, depression among the general population during the COVID-19 pandemic: A systematic review and meta-analysis. *Glob. Health* **2020**, *16*, 57. [\[CrossRef\]](#) [\[PubMed\]](#)
49. Xiao, H.; Carney, D.M.; Youn, S.J.; Janis, R.A.; Castonguay, L.G.; Hayes, J.A.; Locke, B.D. Are we in crisis? National mental health and treatment trends in college counseling centers. *Psychol. Serv.* **2017**, *14*, 407–415. [\[CrossRef\]](#) [\[PubMed\]](#)
50. Bor, W.; Dean, A.J.; Najman, J.; Hayatbakhsh, R. Are child and adolescent mental health problems increasing in the 21st century? A systematic review. *Aust. N. Z. J. Psychiatry* **2014**, *48*, 606–616. [\[CrossRef\]](#) [\[PubMed\]](#)
51. Duffy, M.E.; Twenge, J.M.; Joiner, T.E. Trends in Mood and Anxiety Symptoms and Suicide-Related Outcomes among U.S. Undergraduates, 2007–2018: Evidence From Two National Surveys. *J. Adolesc. Health* **2019**, *65*, 590–598. [\[CrossRef\]](#) [\[PubMed\]](#)
52. Stress in America: Paying with Our Health. American Psychological Association (APA). 2015. Available online: <https://www.apa.org/news/press/releases/stress/2014/stress-report.pdf> (accessed on 23 April 2021).
53. Mental Health Statistics: Stress. Mental Health Foundation. 2018. Available online: <https://www.mentalhealth.org.uk/statistics/mental-health-statistics-stress> (accessed on 23 April 2021).
54. Grados, J.L. Prevalence and prognosis of stress disorders: A review of the epidemiologic literature. *Clin. Epidemiol.* **2017**, *9*, 251–260. [\[CrossRef\]](#) [\[PubMed\]](#)
55. Bandelow, B.; Michaelis, S. Epidemiology of anxiety disorders in the 21st century. *Dialogues Clin. Neurosci.* **2015**, *17*, 327–335. [\[PubMed\]](#)
56. Wang, J.; Wu, X.; Lai, W.; Long, E.; Zhang, X.; Li, W.; Zhu, Y.; Chen, C.; Zhong, X.; Liu, Z.; et al. Prevalence of depression and depressive symptoms among outpatients: A systematic review and meta-analysis. *BMJ Open* **2017**, *7*. [\[CrossRef\]](#) [\[PubMed\]](#)
57. Ritchie, H.; Roser, M. Mental Health. *Our World in Data* 2018. Available online: <https://ourworldindata.org/mental-health> (accessed on 23 April 2021).
58. Panchal, N.; Kamal, R.; Follow, C.C.; Follow, R.G. The Implications of COVID-19 for Mental Health and Substance Use. 2021. Available online: <https://www.kff.org/coronavirus-COVID-19/issue-brief/the-implications-of-COVID-19-for-mental-health-and-substance-use/> (accessed on 23 April 2021).
59. Pierce, M.; Hope, H.; Ford, T.; Hatch, S.; Hotopf, M.; John, A.; Kontopantelis, E.; Webb, R.; Wessely, S.; McManus, S.; et al. Mental health before and during the COVID-19 pandemic: A longitudinal probability sample survey of the UK population. *Lancet Psychiatry* **2020**, *7*, 883–892. [\[CrossRef\]](#)
60. Ettman, C.K.; Abdalla, S.M.; Cohen, G.H.; Sampson, L.; Vivier, P.M.; Galea, S. Prevalence of Depression Symptoms in US Adults Before and During the COVID-19 Pandemic. *JAMA Netw. Open* **2020**, *3*, e2019686. [\[CrossRef\]](#) [\[PubMed\]](#)
61. Jamison, D.; Breman, J.; Measham, A.; Alleyne, G.; Claeson, M.; Evans, D.; Jha, P.; Mills, A.; Musgrove, P. *Disease Control Priorities in Developing Countries*; NCBI Bookshelf; World Bank Publications: Washington, DC, USA, 2006.

62. Wang, P.S.; Aguilar-Gaxiola, S.; Alonso, J.; Angermeyer, M.C.; Borges, G.; Bromet, E.J.; Bruffaerts, R.; de Girolamo, G.; de Graaf, R.; Gureje, O.; et al. Use of mental health services for anxiety, mood, and substance disorders in 17 countries in the WHO world mental health surveys. *Lancet* **2007**, *370*, 841–850. [\[CrossRef\]](#)
63. Schmidtke, A.; Bille-Brahe, U.; DeLeo, D.; Kerkhof, A.; Bjerke, T.; Crepet, P.; Haring, C.; Hawton, K.; Lönnqvist, J.; Michel, K.; et al. Attempted suicide in Europe: Rates, trends and sociodemographic characteristics of suicide attempters during the period 1989–1992. Results of the WHO/EURO Multicentre Study on Parasuicide. *Acta Psychiatr. Scand.* **1996**, *93*, 327–338. [\[CrossRef\]](#)
64. Thornicroft, G.; Chatterji, S.; Evans-Lacko, S.; Gruber, M.; Sampson, N.; Aguilar-Gaxiola, S.; Al-Hamzawi, A.; Alonso, J.; Andrade, L.; Borges, G.; et al. Undertreatment of people with major depressive disorder in 21 countries. *Br. J. Psychiatry* **2017**, *210*, 119–124. [\[CrossRef\]](#)
65. *Investing in Mental Health*; World Health Organization: Geneva, Switzerland, 2003.
66. Winkler, P.; Krupchanka, D.; Roberts, T.; Kondratova, L.; Machů, V.; Höschl, C.; Sartorius, N.; Van Voren, R.; Aizberg, O.; Bitter, I.; et al. A blind spot on the global mental health map: A scoping review of 25 years' development of mental health care for people with severe mental illnesses in central and eastern Europe. *Lancet Psychiatry* **2017**, *4*, 634–642. [\[CrossRef\]](#)
67. *Inequalities in Access to Healthcare*; European Commission: Brussels, Belgium, 2018.
68. Kolenik, T.; Gams, M. Persuasive Technology for Mental Health: One Step Closer to (Mental Health Care) Equality? *IEEE Technol. Soc. Mag.* **2021**, *40*, 80–86. [\[CrossRef\]](#)
69. Angermeyer, M.C.; Matschinger, H. The effect of personal experience with mental illness on the attitude towards individuals suffering from mental disorders. *Soc. Psychiatry Psychiatr. Epidemiol.* **1996**, *31*, 321–326. [\[CrossRef\]](#) [\[PubMed\]](#)
70. Montag, C.; Duke, É.; Markowitz, A. Toward Psychoinformatics: Computer Science Meets Psychology. *Comput. Math. Methods Med.* **2016**, *2016*, 2983685. [\[CrossRef\]](#)
71. Gutierrez, L.J.; Rabbani, K.; Ajayi, O.J.; Gebresilassie, S.K.; Rafferty, J.; Castro, L.A.; Banos, O. Internet of Things for Mental Health: Open Issues in Data Acquisition, Self-Organization, Service Level Agreement, and Identity Management. *Int. J. Environ. Res. Public Health* **2021**, *18*, 1327. [\[CrossRef\]](#) [\[PubMed\]](#)
72. Jain, Y.; Gandhi, H.; Burte, A.; Vora, A. Mental and Physical Health Management System Using ML, Computer Vision and IoT Sensor Network. In Proceedings of the 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 5–7 November 2020; pp. 786–791. [\[CrossRef\]](#)
73. Wikipedia Contributors. Mental Health Informatics—Wikipedia, The Free Encyclopedia. 2021. Available online: [https://en.wikipedia.org/wiki/Mental\\_health\\_informatics](https://en.wikipedia.org/wiki/Mental_health_informatics) (accessed on 24 April 2021).
74. McCrone, P.; Knapp, M.; Proudfoot, J.; Ryden, C.; Cavanagh, K.; Shapiro, D.A.; Ilson, S.; Gray, J.A.; Goldberg, D.; Mann, A.; et al. Cost-effectiveness of computerised cognitive-behavioural therapy for anxiety and depression in primary care: Randomised controlled trial. *Br. J. Psychiatry* **2004**, *185*, 55–62. [\[CrossRef\]](#) [\[PubMed\]](#)
75. Cliffe, B.; Croker, A.; Denne, M.; Stallard, P. Clinicians' use of and attitudes towards technology to provide and support interventions in child and adolescent mental health services. *Child Adolesc. Ment. Health* **2020**, *25*, 95–101. [\[CrossRef\]](#) [\[PubMed\]](#)
76. Mohr, D.C.; Burns, M.N.; Schueller, S.M.; Clarke, G.; Klinkman, M. Behavioral Intervention Technologies: Evidence review and recommendations for future research in mental health. *Gen. Hosp. Psychiatry* **2013**, *35*, 332–338. [\[CrossRef\]](#)
77. Freedman, N.; Hoffenberg, J.D.; Vorus, N.; Frosch, A. The effectiveness of psychoanalytic psychotherapy: The role of treatment duration, frequency of sessions, and the therapeutic relationship. *J. Am. Psychoanal. Assoc.* **1999**, *47*, 741–772. [\[CrossRef\]](#)
78. Sandell, R.; Blomberg, J.; Lazar, A.; Carlsson, J.; Broberg, J.; Schubert, J. Varieties of long-term outcome among patients in psychoanalysis and long-term psychotherapy. A review of findings in the Stockholm Outcome of Psychoanalysis and Psychotherapy Project (STOPP). *Int. J. Psychoanal.* **2000**, *81 Pt 5*, 921–942. [\[CrossRef\]](#)
79. Corrigan, P.; Watson, A. The impact of stigma on people with mental illness. *World Psychiatry Off. J. World Psychiatr. Assoc. (WPA)* **2002**, *1*, 16–20.
80. Amaral, I.; Daniel, F. *Ageism and IT: Social Representations, Exclusion and Citizenship in the Digital Age*; Human Aspects of IT for the Aged Population. Healthy and Active Aging; Zhou, J., Salvendy, G., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 159–166.
81. Pigato, M. *Information and Communication Technology, Poverty, and Development in Sub-Saharan Africa and South Asia*; The World Bank: Washington, DC, USA, 2001; Number 20.
82. Lee, S.G.; Trimi, S.; Kim, C. The impact of cultural differences on technology adoption. *J. World Bus.* **2013**, *48*, 20–29. [\[CrossRef\]](#)
83. *AfriCHI '18: Proceedings of the Second African Conference for Human Computer Interaction: Thriving Communities*; Association for Computing Machinery: New York, NY, USA, 2018.
84. Yorita, A.; Egerton, S.; Chan, C.; Kubota, N. Chatbot for Peer Support Realization based on Mutual Care. In Proceedings of the 2020 IEEE Symposium Series on Computational Intelligence (SSCI), Canberra, Australia, 1–4 December 2020; pp. 1601–1606. [\[CrossRef\]](#)
85. Avancha, S.; Baxi, A.; Kotz, D. Privacy in Mobile Technology for Personal Healthcare. *ACM Comput. Surv.* **2012**, *45*. [\[CrossRef\]](#)
86. Lee, S.S.; Lim, Y.k.; Lee, K.P. A Long-Term Study of User Experience towards Interaction Designs That Support Behavior Change. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems*; Association for Computing Machinery: New York, NY, USA, 2011; pp. 2065–2070. [\[CrossRef\]](#)
87. Eyal, N.; Hoover, R. *Hooked: How to Build Habit-Forming Products*; Penguin Publishing Group: New York, NY, USA, 2014.

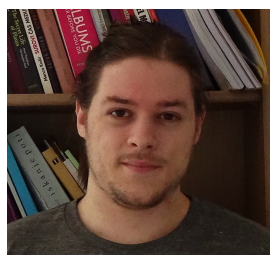


88. Bakker, D.; Kazantzis, N.; Rickwood, D.; Rickard, N. Mental Health Smartphone Apps: Review and Evidence-Based Recommendations for Future Developments. *JMIR Ment. Health* **2016**, *3*, e7. [CrossRef] [PubMed]
89. Chan, S.; Godwin, H.; Gonzalez, A.; Yellowlees, P.M.; Hilty, D.M. Review of Use and Integration of Mobile Apps into Psychiatric Treatments. *Curr. Psychiatry Rep.* **2017**, *19*, 96. [CrossRef]
90. Torous, J.; Nicholas, J.; Larsen, M.E.; Firth, J.; Christensen, H. Clinical review of user engagement with mental health smartphone apps: Evidence, theory and improvements. *Evid. Based Ment. Health* **2018**, *21*, 116–119. [CrossRef]
91. Montenegro, J.L.Z.; da Costa, C.A.; da Rosa Righi, R. Survey of conversational agents in health. *Expert Syst. Appl.* **2019**, *129*, 56–67. [CrossRef]
92. Safi, Z.; Abd-Alrazaq, A.; Khalifa, M.; Househ, M. Technical Aspects of Developing Chatbots for Medical Applications: Scoping Review. *J. Med. Internet Res.* **2020**, *22*, e19127. [CrossRef]
93. Abd-Alrazaq, A.; Safi, Z.; Alajlani, M.; Warren, J.; Househ, M.; Denecke, K. Technical Metrics Used to Evaluate Health Care Chatbots: Scoping Review. *J. Med. Internet Res.* **2020**, *22*, e18301. [CrossRef]
94. Pereira, J.; Díaz, Ó. Using Health Chatbots for Behavior Change: A Mapping Study. *J. Med. Syst.* **2019**, *43*, 135. [CrossRef] [PubMed]
95. Silva, R.; Neiva, F. *Systematic Literature Review in Computer Science—A Practical Guide*; Technical Report; Federal University of Juiz de Fora: Juiz de Fora, Brazil, 2016.
96. Arksey, H.; O'Malley, L. Scoping studies: Towards a methodological framework. *Int. J. Soc. Res. Methodol.* **2005**, *8*, 19–32. [CrossRef]
97. Levac, D.; Colquhoun, H.; O'Brien, K.K. Scoping studies: Advancing the methodology. *Implement. Sci.* **2010**, *5*, 69. [CrossRef]
98. Daudt, H.M.; van Mossel, C.; Scott, S.J. Enhancing the scoping study methodology: A large, inter-professional team's experience with Arksey and O'Malley's framework. *BMC Med. Res. Methodol.* **2013**, *13*, 48. [CrossRef] [PubMed]
99. Petticrew, M.; Roberts, H. *Systematic Reviews in the Social Sciences: A Practical Guide*; Wiley: Hoboken, NJ, USA, 2008.
100. Howland, J.; Wright, T.; Boughan, R.; Roberts, B. How Scholarly Is Google Scholar? A Comparison to Library Databases. *Coll. Res. Libr.* **2009**, *70*, 227–234. [CrossRef]
101. Walters, W. Google Scholar coverage of a multidisciplinary field. *Inf. Process. Manag.* **2007**, *43*, 1121–1132. [CrossRef]
102. Harzing, A.W.; Alakangas, S. Google Scholar, Scopus and the Web of Science: A longitudinal and cross-disciplinary comparison. *Scientometrics* **2016**, *106*, 787–804. [CrossRef]
103. López-Cózar, E.D.; Orduna-Malea, E.; Martín-Martín, A. *Google Scholar as a Data Source for Research Assessment*; Springer: Cham, Switzerland, 2018.
104. Abbott, A. COVID's mental-health toll: How scientists are tracking a surge in depression. *Nature* **2021**, *590*, 194–195. [CrossRef]
105. Medhi Thies, I.; Menon, N.; Magapu, S.; Subramony, M.; O'Neill, J. How Do You Want Your Chatbot? An Exploratory Wizard-of-Oz Study with Young, Urban Indians. In *Human-Computer Interaction—INTERACT 2017*; Bernhaupt, R., Dalvi, G., Joshi, A., K. Balkrishan, D., O'Neill, J., Winckler, M., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp. 441–459.
106. Albright, G.; Adam, C.; Serri, D.; Bleeker, S.; Goldman, R. Harnessing the power of conversations with virtual humans to change health behaviors. *mHealth* **2016**, *2*, 44. [CrossRef]
107. Linchpin. 25 Chatbot Stats and Trends Shaping Businesses in 2021. Available online: <https://linchpinseo.com/chatbot-statistics-trends/> (accessed on 10 May 2021).
108. Yoon, C.; Jeong, C.; Rolland, E. Understanding individual adoption of mobile instant messaging: A multiple perspectives approach. *Inf. Technol. Manag.* **2015**, *16*, 139–151. [CrossRef]
109. Dialogflow Documentation. Available online: <https://cloud.google.com/dialogflow/docs/> (accessed on 23 April 2021).
110. IBM Watson. Available online: <https://www.ibm.com/watson> (accessed on 23 April 2021).
111. Microsoft Bot Framework. Available online: <https://dev.botframework.com/> (accessed on 23 April 2021).
112. GPT-3. Available online: <https://gpt3.website/> (accessed on 23 April 2021).
113. 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.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **2021**, *372*. [CrossRef]
114. Ghandeharioun, A.; McDuff, D.; Czerwinski, M.; Rowan, K. EMMA: An Emotion-Aware Wellbeing Chatbot. In *Proceedings of the 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*, Memphis, TN, USA, 9–12 October 2019; pp. 1–7. [CrossRef]
115. Khadikar, S.; Sharma, P.; Paygude, P. Compassion Driven Conversational Chatbot Aimed for better Mental Health. *Zeich. J.* **2020**, *6*, 121–127.
116. Park, S.; Choi, J.; Lee, S.; Oh, C.; Kim, C.; La, S.; Lee, J.; Suh, B. Designing a Chatbot for a Brief Motivational Interview on Stress Management: Qualitative Case Study. *J. Med. Internet Res.* **2019**, *21*, e12231. [CrossRef] [PubMed]
117. Pola, S.; Sheela Rani Chetty, M. Behavioral therapy using conversational chatbot for depression treatment using advanced RNN and pretrained word embeddings. *Mater. Today Proc.* **2021**. [CrossRef]
118. Rishabh, C.; Anuradha, J. Counsellor chatbot. *Int. Res. J. Comput. Sci.* **2018**, *3*, 126–136.
119. Wikipedia Contributors. Artificial Linguistic Internet Computer Entity—Wikipedia, The Free Encyclopedia. 2020. Available online: [https://en.wikipedia.org/wiki/Artificial\\_Linguistic\\_Internet\\_Computer\\_Entity](https://en.wikipedia.org/wiki/Artificial_Linguistic_Internet_Computer_Entity) (accessed on 24 April 2021).



120. Yorita, A.; Egerton, S.; Oakman, J.; Chan, C.; Kubota, N. A Robot Assisted Stress Management Framework: Using Conversation to Measure Occupational Stress. In Proceedings of the 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Miyazaki, Japan, 7–10 October 2018; pp. 3761–3767.
121. Gould, C.E.; Ma, F.; Loup, J.R.; Juang, C.; Sakai, E.Y.; Pepin, R. Technology-based mental health assessment and intervention. In *Handbook of Mental Health and Aging*; Academic Press: New York, NY, USA, 2020; pp. 401–415. [\[CrossRef\]](#)
122. Kolenik, T. Methods in Digital Mental Health: Smartphone-based Assessment and Intervention for Stress, Anxiety and Depression. In *Integrating Artificial Intelligence and IoT for Advanced Health Informatics*; Comito, C., Forestiero, A., Zumpano, E., Eds.; Springer: Berlin/Heidelberg, Germany, 2021. In press.
123. Bocklisch, T.; Faulkner, J.; Pawlowski, N.; Nichol, A. Rasa: Open Source Language Understanding and Dialogue Management. *arXiv* **2017**, arXiv:1712.05181.
124. Ghandeharioun, A.; McDuff, D.; Czerwinski, M.; Rowan, K. Towards Understanding Emotional Intelligence for Behavior Change Chatbots. In Proceedings of the 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII), Memphis, TN, USA, 9–12 October 2019; pp. 8–14. [\[CrossRef\]](#)
125. Daws, R. Medical Chatbot Using OpenAI's GPT-3 Told a Fake Patient to Kill Themselves. 2020. Available online: <https://artificialintelligence-news.com/2020/10/28/medical-chatbot-openai-gpt3-patient-kill-themselves/> (accessed on 23 April 2021).
126. Suganuma, S.; Sakamoto, D.; Shimoyama, H. An Embodied Conversational Agent for Unguided Internet-Based Cognitive Behavior Therapy in Preventative Mental Health: Feasibility and Acceptability Pilot Trial. *JMIR Ment. Health* **2018**, *5*, e10454. [\[CrossRef\]](#) [\[PubMed\]](#)
127. Oyibo, K.; Orji, R.; Vassileva, J. Investigation of the Influence of Personality Traits on Cialdini's Persuasive Strategies. PPT@PERSUASIVE. 2017. Available online: [http://ceur-ws.org/Vol-1833/4\\_Oyibo.pdf](http://ceur-ws.org/Vol-1833/4_Oyibo.pdf) (accessed on 23 April 2021).
128. Kaity, M.; Balakrishnan, V. Sentiment lexicons and non-English languages: A survey. *Knowl. Inf. Syst.* **2020**, *62*. [\[CrossRef\]](#)

## Short Biography of Authors



**Tine Kolenik** Tine Kolenik is a Junior Researcher at the Department of Intelligent Systems at the Jožef Stefan Institute and an Assistant Lecturer at the Jožef Stefan International Postgraduate School. He is primarily interested in (idiographic) change, investigated on quantitative and qualitative levels of analysis through cognitive science, artificial intelligence, persuasive technology, attitude and behavior change support systems, digital mental health, behavioral data science, natural and artificial cognitive architectures, philosophy of science, and 5E cognition. His PhD is focused on the intelligent cognitive assistant technology for attitude and behavior change in mental health. Kolenik is also an assistant editor for the international journal of computing and informatics *Informatica* and Center for Cognitive Science of University of Ljubljana collaborator.



**Matjaž Gams** Prof. Dr. Matjaž Gams is head of the Department of Intelligent Systems at the Jožef Stefan Institute and professor of computer and information science at the University of Ljubljana and Jožef Stefan International Postgraduate School. He has been a member of multiple national councils, namely the National Council for Science and Technology, and is currently a national councillor for research activity. He is a member of programme committees of 10–20 international conferences annually, including IJCAI and AAAI—two of the most prestigious conferences in artificial intelligence, as well as a member of PO of 12 international publications. Additionally, he has been a contact executive editor of journal *Informatica*.