



# Recommendations for the ethical use and design of artificial intelligent care providers



David D. Luxton<sup>a,b,\*</sup>

<sup>a</sup> Department of Psychiatry and Behavioral Sciences, University of Washington School of Medicine, Box 356560, Seattle, WA 98195, United States

<sup>b</sup> National Center for Telehealth & Technology (T2), 9933 West Hayes Street, Madigan Annex, Joint Base Lewis-McChord, Tacoma, WA 98431, United States

## ARTICLE INFO

### Article history:

Received 27 March 2014

Received in revised form 30 May 2014

Accepted 24 June 2014

### Keywords:

Artificial intelligent agents

Ethics

Ethical codes

Practice guidelines

Care providers

Mental health

## ABSTRACT

**Objective:** This paper identifies and reviews ethical issues associated with artificial intelligent care providers (AICPs) in mental health care and other helping professions. Specific recommendations are made for the development of ethical codes, guidelines, and the design of AICPs.

**Methods:** Current developments in the application of AICPs and associated technologies are reviewed and a foundational overview of applicable ethical principles in mental health care is provided. Emerging ethical issues regarding the use of AICPs are then reviewed in detail. Recommendations for ethical codes and guidelines as well as for the development of semi-autonomous and autonomous AICP systems are described. The benefits of AICPs and implications for the helping professions are discussed in order to weigh the pros and cons of their use.

**Results:** Existing ethics codes and practice guidelines do not presently consider the current or the future use of interactive artificial intelligent agents to assist and to potentially replace mental health care professionals. AICPs present new ethical issues that will have significant ramifications for the mental health care and other helping professions. Primary issues involve the therapeutic relationship, competence, liability, trust, privacy, and patient safety. Many of the same ethical and philosophical considerations are applicable to use and design of AICPs in medicine, nursing, social work, education, and ministry.

**Conclusion:** The ethical and moral aspects regarding the use of AICP systems must be well thought-out today as this will help to guide the use and development of these systems in the future. Topics presented are relevant to end users, AI developers, and researchers, as well as policy makers and regulatory boards.

Published by Elsevier B.V.

## 1. Introduction

Nearly half a century ago Joseph Weizenbaum introduced ELIZA, the first simulation of a psychotherapist [1]. ELIZA, also known as DOCTOR, was a simple computer program that was capable of mimicking the question and response conversation of a psychotherapeutic interview. A few years later, psychiatrist Kenneth Colby developed a program called PARRY that simulated a person with paranoid schizophrenia [2]. Advancements in artificial intelligence (AI) and associated technologies, such as virtual reality, natural language processing, and affective computing have enabled the creation of artificial intelligent agents in the form

of highly realistic simulated psychotherapists, counselors, and therapeutic coaches. These modern systems, which may be considered to be the conceptual evolution of primitive “chatterbot” systems such as ELIZA and PARRY, are capable of carrying on highly interactive and intelligent conversations and can be used to provide counseling, training, clinical assessment, and other therapeutic functions [3].

The practice of mental health care entails significant ethical responsibilities that involve consideration of complex legal, moral, cultural, and personal factors. The professions of psychology, counseling, and psychiatry, for example, all have ethical codes of conduct that help guide ethical decision making and behavior of care providers; however, existing professional ethics codes and practice guidelines do not presently consider the current or the future use of artificial intelligent agents to assist or potentially replace humans in these professions. As history has demonstrated, rapidly changing technology can get ahead of the awareness of the greater population and thus laws and guidelines have to catch up with technology. My goal with this paper is thus to discuss emerging ethical issues

\* Correspondence to: National Center for Telehealth and Technology (T2), Defense Centers of Excellence for Psychological Health and Traumatic Brain Injury, 9933 West Hayes Street, Madigan Annex, Joint Base Lewis-McChord, WA 98431, United States. Tel.: +1 253 968 3581.

E-mail address: [ddluxton@uw.edu](mailto:ddluxton@uw.edu)

associated with artificial intelligent agents that are designed to provide interactive mental health care services (i.e., psychotherapy, counseling, clinical assessment, etc.). Many of the same ethical and philosophical considerations are also applicable to the use of this technology in other helping professions such as medicine, nursing, social work, education, and ministry.

I begin by providing an overview of current developments in artificial intelligent care providers (AICPs) in order to illustrate current capabilities and future directions of the technology. I also present a brief overview of professional ethics codes in order to orient the reader to the overarching values and ethical principles of the mental health care disciplines. I do not survey the depth of ethical theory or the technical aspects of the design of artificial moral agents as these are covered elsewhere [4–7]. Rather, I focus on practical ethical issues and what may be needed in future professional ethics codes, laws, and practice guidelines. I also discuss the pros and cons regarding the use of AICPs in the mental health care and other helping professions. The topics that I present are not only important for being ethical users of these technologies, but for the design of these technologies in order to make them effective at what they are intended to do. Thus, what I discuss here should be of interest to end users, developers, and researchers as well as professional organizations and regulatory boards in the years ahead.

## 2. Current state of the art

AICPs can be designed in various forms to interact with users including virtual reality simulations (avatars), robots (humanoid or non-humanoid), or non-embodied systems that consist of only voice simulation and environmental sensors. A leading area of development of AICPs is the creation of virtual human avatars that make use of advancements in virtual reality simulation, natural language processing, and knowledge-based artificial intelligence. Life-like virtual humans have been developed and tested for use in clinical training and skill acquisition and to provide care seekers with information about mental health resources and support [8,9]. SimCoach ([www.simcoach.org](http://www.simcoach.org)), for example, is an online avatar-based virtual intelligent agent system that is designed to interact with and connect military service members and their families to health care and other helping resources [9]. Virtual intelligent agent systems have also been developed and tested to help medication adherence among patients with schizophrenia [10] and to provide patients with hospital discharge planning [11].

Current developments in affective sensing and processing are providing artificial intelligent systems with capabilities to detect, interpret, and express emotions as well as detect other behavioral signals of humans that they interact with. For example, the Defense Advanced Research Projects Agency (DARPA) Detection and Computational Analysis of Psychological Signals (DCAPS) system uses machine learning, natural language processing, and computer vision to analyze language, physical gestures, and social signals to detect psychological distress cues in humans [12]. The intent of the system is to help improve the psychological health of military personnel as well as to develop and test algorithms for detecting distress markers in humans from various inputs including sleep patterns, voice and data communications, social interactions, Internet use behaviors, and non-verbal cues (e.g., facial gestures and body posture and movement).

Robotics and other intelligent technologies are also advancing at an incredible pace and are finding very practical applications in the field of medical care. Robots that can interact with patients and medical staff, such as RP-VITA [13], are being tested and deployed in hospitals. IBM has developed an expanded, commercially available version of the natural language processing DeepQA system *Watson* that has learned the medical literature, therefore

allowing it to serve as an interactive medical knowledge expert and consultant [14,15]. These emerging technologies have the potential to greatly expand the capabilities of AICPs through integration of them. In the future, it may be possible to build what I call the “super clinician” [3], an AI system that integrates optical and auditory sensors, natural language processing, knowledge-based AI, and non-verbal behavior detection that could be used to conduct psychotherapy and other clinical functions. Computer vision processing could be used to observe facial expressions, speech analysis technologies could assess inflection and tone of voice, and natural language processing could be used to detect semantic representation of emotional states. The use of infra-red sensors, for example, could enable such a system to detect physiological processes that are undetectable by humans and will make it far superior to the capabilities of human clinicians.

The technological capabilities of AICPs are rapidly expanding and further public and private investment in AICP development and application can be expected. The practical applications of AICPs include screening, assessments, and counseling for self-care as well as for traditional clinical care in both government and private health care institutions. Other practical applications include use in corporate employee assistance programs, in prisons for forensic assessments and evaluations, and in austere or remote environments where human care providers are few, such as in submarines or during orbital or interplanetary space travel. Just as with so many other new technologies (e.g., Apple’s Siri), we can expect AICPs to become a ubiquitous aspect of our society in the years ahead.

## 3. Current applicable ethics codes and guidelines

### 3.1. Overview of professional codes of ethics

The American Psychiatric Association (APA), American Psychological Association (APA), and the American Counseling Association (ACA) are examples of the largest mental healthcare professional organizations in the United States that have published ethical codes for their respective disciplines. There are also several national certification boards, state regulatory boards, and specialty areas that have their own ethics or professional practice guidelines [16]. The specific guidelines of these professional organizations and boards cover the range of areas applicable to mental health care practice including providing treatments, clinical assessments, research, training and consultation. In general, ethical codes and guidelines are intended to help guide the behavior of health care professionals and organizations toward the benefit and protection of patients. Ethical codes and guidelines also help practitioners to resolve ethical dilemmas and to justify their decisions and courses of action. Further, they protect care providers and their institutions by setting standards of conduct that ultimately promote the trust of patients, professional colleagues, and the general public.

While there are differences between the ethical codes and guidelines of the various mental healthcare organizations, there are several key common themes. As noted by Koocher and Keith-Speigel [17], these include:

1. Promoting the welfare of consumers (patients)
2. Practicing within scope of one’s competence
3. Doing no harm (non-maleficence)
4. Protecting the patients’ confidentiality and privacy
5. Acting ethically and responsibly
6. Avoiding exploitation
7. Upholding integrity of the profession by striving for aspirational practice

**Table 1**  
Engineering and Physical Sciences Research Council (EPSRC)/Arts and Humanities Research Council (AHRC) Ethical principles regarding robots.

1. Robots should not be designed solely or primarily to kill or harm humans.
2. Humans, not robots, are responsible agents. Robots are tools designed to achieve human goals.
3. Robots should be designed in ways that assure their safety and security.
4. Robots are artifacts; they should not be designed to exploit vulnerable users by evoking an emotional response or dependency. It should always be possible to tell a robot from a human.
5. It should always be possible to find out who is legally responsible for a robot.

Professional ethics codes typically include both aspirational principles (e.g., beneficence, non-maleficence, autonomy, dignity and integrity) as well as standards that are enforceable rules. Guidelines are not ethical codes or legal standards; however, provisions within may reflect the spirit of the overarching medical ethical principles and help to assure competence of practitioners. Laws are enforceable obligations and in some cases, laws may be in conflict with ethical responsibilities. The American Psychological Association addresses this with the following provision: “If psychologists’ ethical responsibilities conflict with law, regulations, or other governing legal authority, psychologists make known their commitment to the Ethics Code and take steps to resolve the conflict. If the conflict is unresolvable via such means, psychologists may adhere to the requirements of the law, regulations, or other governing authority” [18]. Several ethical codes, such as that of the American Psychiatric Association, specify that care providers (physicians) should work to change the law if they believe that the law is unjust [19].

### 3.2. Technology specific ethical codes and guidelines

In recognizing the need to address ethical issues and best practices that arise with new technological advancements, several mental health care professional organizations have included provisions in their ethical guidelines or as supplements that address topics such as electronic data security, use of the Internet for providing care services, or the use of electronic communication with patients. While some of the existing mental health care ethics codes (e.g., the American Psychological Association) provide general provisions for competence of care in emerging areas, these provisions are entirely general and are not sufficient to address the emerging ethical implications associated with artificial intelligent systems as care providers. Other organizations, such as the National Board of Certified Counselors, The International Society for Mental Health, the American Medical Association, and the American Telemedicine Association have established specific guidelines regarding the use of the Internet to provide care; however, these do not yet consider the application of AICPs to provide care services.

Topics discussed in the field of machine ethics [4] and associated fields are of high relevance to the ethical design and use of AICPs. Machine ethics is concerned with the moral behavior of artificial intelligent agents. Artificial moral agents (AMAs) are robots or artificial intelligent systems that behave toward human users or other machines in an ethical and moral fashion [4]. Roboethics [20,21] is also an emerging field that is concerned with the ethical behavior of humans when it comes to designing, creating, and interacting with artificial intelligent entities (i.e., robots). Science fiction has also approached the topic of ethics with intelligent machines, most notably Isaac Asimov’s Three laws (and one additional law) of Robotics [22,23]. In 2011, the Engineering and Physical Sciences Research Council (EPSRC) and the Arts and Humanities Research Council (AHRC) (Great Britain) published a set of ethical principles for designers, builders and users of robots. These are shown in Table 1. Riek and Howard [24] have also discussed emerging ethical,

design, and legal considerations associated with human–robot interaction.

While ethics for humans users of artificial intelligent agents can be conceptualized as distinct from the ethics of intelligent machines [4], both are highly relevant to the use of AICPs, especially as these systems are designed to be increasingly autonomous. According to Moor [25], *implicit ethical agents* are machines constrained by what programmers will have them do whereas *explicit ethical agents* are machines that calculate what is ethical on their own by applying ethical principles to diverse and varying degrees of complex situations. Current virtual avatar systems, for example, can be considered to be implicit ethical agents; however, as they become more autonomous, they will need to function as explicit ethical agents. Consider an artificial intelligent agent that is capable of independently performing diagnostic assessments, providing treatment, and monitoring the symptoms of a depressed patient. The system will need to be able to make decisions and select courses of action that are consistent with applicable ethics codes during its interaction with patients and also be capable of resolving complex ethical dilemmas, just as a human psychotherapist is required to do. Thus, the relevance of professional ethics and machine ethics can be expected to increase as AICPs are put to use in the mental health care and other helping fields.

## 4. Ethical issues associated with artificially intelligent care providers

### 4.1. Relationship between care providers and care seekers

The *therapeutic relationship*, also called the *working alliance*, is a term used to describe the professional relationship between a provider of care (i.e., a psychotherapist) and a patient whereby the care provider hopes to engage with and achieve therapeutic goals with a patient [26]. The therapeutic relationship is a key common factor associated with desirable treatment outcomes that is independent of the specific type of treatment. In other words, the quality of the therapeutic relationship accounts for why clients improve (or fail to improve) at least as much as the particular treatment method [27,28].

The relationship between professional care providers and patients is not an ordinary social relationship; there are important legal and ethical obligations because care providers are in a position of power in their relationship with patients and there is potential for harm and exploitation of patients. The ethical obligations of the relationship between care providers and patients are central to the ethical codes of the mental health care professions. The American Psychiatric Association Ethical Principles states:

“The psychiatrist shall be ever vigilant about the impact that his or her conduct has upon the boundaries of the doctor–patient relationship, and thus upon the well-being of the patient. These requirements become particularly important because of the essentially private, highly personal, and sometimes intensely emotional nature of the relationship established with the psychiatrist.” [29]

Central to the therapeutic relationship is trust as this forms the basis for the care provider–patient relationship. In order to make accurate diagnoses and provide optimal treatments, the patient must be able to feel comfortable to communicate all relevant clinical information to care providers. Just as between human care providers and their patients, trust will be a critical variable for patients who interact with AICPs.

The issue of trust raises an important philosophical question; is it unethical to simulate a human so much that people believe

that the simulation is human? Weizenbaum [1] notes that ELIZA did indeed deceive people:

“Nevertheless, ELIZA created the most remarkable illusion of having understood in the minds of the many people who conversed with it. People who knew very well that they were conversing with a machine soon forgot that fact, just as theatergoers, in the grip of suspended disbelief, soon forget that the action they are witnessing is not “real”. This illusion was especially strong and most tenaciously clung to among people who knew little or nothing about computers. They would often demand to be permitted to converse with the system in private, and would, after conversing with it for a time, insist, in spite of my explanations, that the machine really understood them” (p. 189).

Weizenbaum observed that people conversing with ELIZA were more than simply suspending belief. He noted that even while people know that they are conversing with a software program, they still believe that the simulated psychotherapist is real because the person assigns meaning and interpretation to what ELIZA says, which “confirms the person’s hypothesis” that the system understands them. We also know from research that people can form strong emotional attachments to technological devices and simulations, such as virtual avatars [30]. We can thus expect persons to confide in AICPs, to experience them as “real persons”, and to form strong attachments to them.

Key to forming and maintaining the therapeutic relationship is *empathetic understanding* [31]. Empathetic understanding conveys to the patient that the care provider understands and cares what the patient is feeling and experiencing. *Reflection* describes the process whereby the care provider communicates that she/he is feeling and responding to the feelings of the patient [31]. ELIZA demonstrated with its basic text interface and few hundred lines of computer code that it is easy to reflect the feelings of a person, espouse empathy, and “pull in” a person interacting with it. Current systems are much more advanced in this regard; simulated care providers can verbally echo important statements expressed by the patient, say “uh ha”, and make facial or body movements that reflect attentiveness and understanding of what a patient is saying or expressing with body movement. Thus, modern simulations have the appearance of empathetic concern that is incredibly interactive and human-like.

Even when a patient is consciously aware that a care provider simulation is a machine, the patient can be expected to experience intense emotions during the interaction and toward the simulation. Emotional engagement (and even intense emotional interaction) may be desirable in the appropriate therapeutic context. For example, *transference* is the process in which patients unconsciously redirect feelings toward one object (a significant person in the patient’s past or current life) to another (i.e., the therapist) [27]. While transference may be desirable and have therapeutic benefits in some contexts, it is important to consider how this may be undesirable as well. For example, some patients may become overly attached or even “attracted” to an AICP in a non-therapeutic manner. A potential for iatrogenic effects of AICPs can be paralleled to the same types of ethical issues and risks identified when using clinical virtual reality including exacerbating or inducing mental health symptoms that were previously not present [32]. Even when disclosure is made that the system is “just a machine”, some patients may believe that the machine is “alive” or that there is a person, or some other malevolent force behind the simulation. Miller [33] and Riek and Watson [34] have discussed the concerns with “Turing Deceptions”, whereby a person is unable to determine if they are interacting with a machine or not. This issue poses an ethical problem, especially when working with vulnerable persons. For example, simulations may be contraindicated for use with persons who have delusional or psychotic

psychopathologies in the absence of careful patient screening and monitoring.

It is a necessary ethical obligation to establish a safe therapeutic environment for patients to experience emotions and for psychotherapists to monitor emotional reactions and attenuate emotional interactions in a clinically appropriate manner. Psychotherapists must also pay special attention to the effects that the end of the therapeutic relationship will have on their patients. AICP systems must be designed to appropriately end its relationship with patients in a manner that does not cause distress or is harmful to patients (i.e., not making appropriate referrals for additional care or addressing the emotional impact of the closure of the therapeutic relationship, etc.). Unsupervised AICPs that are unable to detect worsening of symptoms would present particular risk.

#### 4.2. Competence and scope of use

Mental health care professionals are responsible for many tasks that require general and specific training and skill. Competence refers to their ability to appropriately perform these tasks. Competence is an ethical obligation because providing services outside of the boundaries of one’s trained skill or expertise could put patients at risk of harm. Moreover, incompetence of professional care providers is damaging to the perception of the mental health care profession when examples of inappropriate or harmful behavior are made public. The ethical codes of the American Psychological Association, American Psychiatric Association and others have specific provisions that address professional competence. The American Psychiatric Association code states “A psychiatrist who regularly practices outside his or her area of professional competence should be considered unethical. Determination of professional competence should be made by peer review boards or other appropriate bodies.” [29]. In the mental healthcare fields, competence is generally maintained through formal education, licensure, specialty training, and continuing education.

The topic of competence is germane to both the design and ethical use of AICPs. Just as with the use of any other technology or technique, mental healthcare professionals (and health care organizations) must be competent in their application of AICP systems. This may include knowing when they are appropriate for use, and what the limitations are, and what the risks are. Moreover, AICP systems must be designed and capable of performing tasks in a competent manner. This includes core competence in application of interpersonal techniques, treatment protocols, safety protocols (how to manage situations where care seekers indicate intent to self-harm or harm another person), and cultural competencies.

A central ethical problem may be the application of AICP systems when they are not adequately controlled based on the scope of their tested capabilities. Systems deployed for access on the Internet that claim to provide particular clinical services or benefits to patients when they are not adequate or appropriate to do the stated services is one example. This is already known to be a problem with unregulated online mental health services and has led to lawsuits in the past [35]. This issue raises a related question; what should be the requirements for showing the credentials of AICPs? The American Psychological Association Ethics Code explicitly states: “Psychologists make available their name, qualifications, registration number, and an indication of where these details can be verified” [16]. The lack of display of the credentials and qualifications of AICP systems and their operators could be damaging to end users and the mental health care profession.

#### 4.3. Accountability and liability risks

The logical extension of discussions regarding scope of use and competence of AICPs and their operators is the topic of liability.

Some of the key questions that arise are: Who should be responsible for the actions, decisions, and recommendations that AICPs make? What if a care seeker dies by suicide or engages in homicide after disclosing intent to an AICP? What is the appropriate responsibility of end users, developers and others as these systems increase in autonomy? Ethics codes, guidelines, and laws will need to consider these types of questions.

A central ethical responsibility of professional health care providers is to assure the safety of patients while they are under care. Take, for example, a patient who discloses intent to end their own life while under care. It is the responsibility of the care provider to take appropriate actions (e.g., in some cases hospitalize the patient, notify emergency services) to assure the safety of the patient. Duty-to-warn statutes require healthcare professionals to notify authorities when a patient makes a threat to another person. In either of these examples, this requires monitoring of content of what the patient discloses, assessing risk factors, and applying clinical judgment (e.g., assessing intent, weighing clinical and other factors) to make decisions regarding the best course of action. The laws regarding duty-to-warn and other requirements may also vary from one jurisdiction to another (i.e., state to state in the United States). Thus, processes must be in place in order to assure that patient safety and that other legal and ethical obligations are met. However, current systems, such as virtual avatars that are accessible via the Internet, may provide services across state lines and for that matter, international boundaries. The issue of practice across jurisdictional boundaries is a major discussion in the field of telemedicine (providing health services over a distance using communication technologies) [35]. Unregulated deployment of these systems, such as on the Internet, will undoubtedly present significant ethical and legal implications if expectations for competence and patient safety are not enforced.

The complexity of liability increases with the use of highly autonomous AICP systems. As noted by Sullins [36], one theoretical model to address liability is the standard user, tool, and victim model. In this model, technology mediates the moral situation between the actor (person who uses the technology) and the victim. When something goes awry, the user is seen as at fault, not the tool. However, this simplistic model does not describe the complexity of when the tool (i.e., an AICP) has a level of moral agency (moral decision making) or when there are other key entities involved, such as the programmers of the technology. Sullins proposes that intelligent machines (i.e., robots) are indeed moral agents when there is reasonable level of abstraction under which the machine has autonomous intentions and responsibilities. If the machine is seen as autonomous, then the machine can be considered a moral agent. If this is the case then the machine may also be seen as responsible, at least in part, for the actions it makes. Furthermore, as the computational complexity of AICPs increase, it will become increasingly difficult to predict how an AICP system will behave in all situations. If in a court of law, human mental health care professionals (or their employers) must demonstrate that their actions were reasonable and consistent with what is typically expected by others in the same profession (i.e., a group of peers), ethical codes, laws and guidelines. If an AICP system exists as a mysterious black box, it may be considered unethical to use it without the capability to demonstrate how the system derived its decisions for a particular course of action.

#### 4.4. Respect of privacy and trust

Respect of privacy reflects the right to autonomy and dignity and it is seen as essential to individual wellbeing [37]. Betrayal of trust due to privacy invasion or unauthorized misuse of information damages the trust of both individual care providers and their profession. Threats to privacy can come in the form of

inappropriate use of private data as well as data security issues including unauthorized access to electronic data.

While threats to patient privacy (electronic data security) are common to many types of technologies used in the practice and management of healthcare [35], current and emerging technological capabilities, such as psychological signal detection (e.g., via visual, voice, psychological data collection), access to “big data”, as well as conversation logs between patients and AICPs create the potential for much more data to be collected about individuals, and without individuals being aware of the collection. Patients can be expected to reveal deeply personal information to AICP systems, of which can cause significant harm to individuals if used inappropriately. The issue of privacy during use of a simulated psychotherapist was highlighted by Weizenbaum [1] who noted that when he informed his secretary (who had been using ELIZA) that he had access to the logs of all the conversations, she reacted with outrage at this invasion of her privacy. Weizenbaum was surprised to find that such a simple program could so easily deceive a naive user into revealing personal information.

Information AICPs may share with human care providers (e.g., to assure continuity of care), as well as potential inconsistency between the information provided by human care providers and AICPs, could result in patient distrust in care providers and be damaging to the therapeutic relationship. Furthermore, the dual-use aspects of the technology may be a threat to the trust that people have of these systems and the helping professions. The fact that the same technologies may be used for multiple purposes other than their stated use in health care may influence the public’s perception of the use of those technologies and thus increase apprehension to participate in activities in which their data are being collected. For example, it is feasible that the same psychological signal detections systems with applicability for mental health care purposes can be used for prisoner interrogation purposes. Also, the possibility that data collected while interacting with AICPs may also be used by corporations, governments or other entities for making predictions about a person’s behavior that have implications, such as for future employment, may also increase distrust and apprehension regarding their use.

The implications regarding data privacy and trust extend beyond the helping fields and have much grander moral ramifications. These technologies present an inherent risk of the perception of and the actual loss of private thoughts. It is already a known fact that current technologies that make use of artificial intelligence are used to track and analyze phone calls, observe and identify people through cameras in public places, and all form of electronic communication in the world [38]. Abuse of this type of personal data that could be collected by AICPs could allow governments or other entities to control individuals or suppress dissent. Hackers also pose a risk to privacy and could potentially exploit very sensitive data of individuals who interact with AICPs. Both the actual and perception of loss of private information are inimical to the fields of mental health practice and to a free society as a whole.

## 5. Recommendations

### 5.1. Requirements for ethical codes and guidelines

I provide a summarized list of recommendations for ethical codes and guidelines in Table 2. To begin, existing professional ethical codes and guidelines will need to be expanded to address the therapeutic relationship between AICPs and patients. Provisions that include requirements for human supervision of AICPs whereby clinicians are responsible for monitoring and addressing therapeutic relationship issues, emotional reactions, and adverse patient reactions that may pose a risk to patient safety should be

**Table 2**  
Summary of considerations and recommendations for ethical codes, guidelines, and use of artificial intelligent care provider (AICP) systems.

1. Ensure appropriateness of the use of AICPs for intended clinical populations and the individual needs of patients
2. Disclose what services and limits of services will be provided by AICPs
3. Ensure that system users understand the capabilities, scope of use, and limitations of AICP systems and describe these to patients as part of the informed consent process
4. Require human supervision and monitoring for adverse patient reactions and clinical contraindications when these may be a risk
5. Display credentials and qualifications of both AICPs and care provider users
6. Describe and disclose data use including extent of data collection, access limitations, and disposition of data records
7. Follow applicable privacy laws and rules
8. Ensure that AICP systems capabilities follow the latest established clinical best practices
9. Ensure that the end of patient–AICP interaction provides continuity of care and sensitivity to the emotional nature of patient–AICPs interactions
10. Provide mechanism for patients to ask questions and report complaints

considered. Requirements for supervision or monitoring should depend on the context and clinical application of AICP systems. For example, AICPs that are used for simple assessment of symptoms, educational coaching, and training purposes may not require the level of supervision or monitoring that more intensive clinical interactions and treatments would require. Provisions that address the competence of users of these systems (i.e., mental health care providers, companies that provide AICP services) are also recommended. These should include requirements to demonstrate that system users understand the capabilities, scope of use, and limitations of AICP systems. For example, psychologists who use these systems in the context of psychological work should have training and knowledge regarding the use of AICPs and in the domain that the AICP is to be used (i.e., psychotherapy, clinical assessment, etc.). There should also be a requirement for AICP systems to be updated to keep current with the latest clinical best practices just as human providers must do.

For AICP systems that are accessed remotely (i.e., via the Internet), we should expect adherence to the same guidelines for assuring patient safety and overall clinical best practices that are used in the field of telemedicine. These include; assessment of whether remote care is appropriate to begin with, assessment of safety risks, the gathering of information regarding emergency services at the patient's location, and involvement of third parties who may assist during emergencies [39]. For instance, if a system is designed to conduct clinical diagnostic assessments, it will need to demonstrate the ability to detect and assess risk for suicide and then take appropriate steps to alert supervisors or administrators to become involved to assure the safety of the patient. Specifications that use of AICP systems must be consistent with applicable laws (e.g., practice jurisdictions, etc.) and established best practices for remote clinical services should be enforced. Legal restrictions on the use of AICPs when presented to the public in untested and uncontrolled manner (such as on the Internet) may also be required. Furthermore, the honest and obvious display of credentials and representation of qualifications and competencies of AICPs and their operators should help consumers make informed decisions and to help protect them from harm resulting from incompetence or unauthorized practice. Patients should also have a way to voice concerns regarding patient safety or quality of services provided by AICPs and to have those issues reviewed and resolved.

Requirements for full disclosure of what the system(s) does, its purpose, its scope, and how people use the information from it needs to be provided to end users. This disclosure should also describe whether the system(s) meets requirements of applicable privacy laws. Rules that specify requirements for maintenance and destruction of AICP data records, including conversation logs,

**Table 3**  
Summary of recommendations for the design of artificial intelligent care provider (AICP) systems.

1. Design to follow the appropriate ethical codes and guidelines consistent with domain of use
2. Identify and provide specifications of use and limits of autonomy of AICP systems to end users
3. Test safety and ethical decision making of AICP systems in diverse ethical situations encountered in all applicable clinical contexts
4. Include capability for data logs/audit trails to track and explain AICP decision making
5. Provide built-in safeguards to assure that systems are only able to provide services within established boundaries of competence and domain of use
6. Consider level of human realism of AICPs including appearance and other behavioral characteristics as appropriate for intended application
7. Consider cultural sensitivity and diversity in design of AICPs

should be developed and disclosed to end users. The disclosure process (i.e., informed consent) should also include clear description of the system and intentions of its use.

Given the potential for fully autonomous intelligent agents in the helping professions, it is conceivable that ethical codes, guidelines, and laws that hold autonomous intelligent systems accountable will be needed. However, AICPs are fundamentally different from human care providers because they cannot accept appropriate responsibility for their actions nor do they have the moral consequences that humans do. Machines will not experience embarrassment, stress, or the pain associated with loss of clinical privileges at a care institution, expulsion from professional organizations, loss of their license to practice, or other legal sanctions. While this topic is open to philosophical debate, it may make most practical sense for legal liability to be explicitly linked to humans.

### 5.2. Recommendations for AICP design and testing

A list of considerations and recommendations for the design of AICPs are shown in Table 3. AICP systems could be built to do work in nearly any combination of domains (psychological work, medicine, counseling, social work, etc.), thus, systems will need to meet the ethical requirements of those respective professional domains based on what activity they are providing. For example, an AICP that functions as a psychiatrist by performing clinical diagnostics and monitoring medication use of patients should follow the ethical guidelines applicable to the field of psychiatry. The developers of AICP systems will need to be able to specify the capabilities and limitations of these systems based on what services they are designed to provide.

How AICPs are designed to recall information when interacting with care seekers is another important issue. Richards and Bransky [40] found that virtual intelligent agents (in the form of real estate agent characters) that were programmed to be poor at recalling information were frustrating to end users, whereas virtual characters that exhibited forgetting by either explicitly acknowledging that they have forgotten something or by not stating it at all were perceived to have more natural memory by end users. While simulation of forgetting and recall facilitated a more believable and natural interaction between the virtual intelligent agent and the end user, it also influenced how much the end user experienced trust in the virtual intelligent agent. This design characteristic is important to consider because of how it may influence the ability for AICPs to effectively interact and engage with patients to achieve therapeutic goals.

Autonomous AICP systems that are intended to fully replace human care providers will have to have the capabilities to assess and monitor emotions just as human care providers do, including the observation of very subtle characteristics of emotional states. Technology is rapidly advancing in the area of affective sensing

and thus the capabilities required for this are becoming feasible. However, ethical codes and guidelines will need to specify requirements for end users to know what the limitations are and to address them in a competent manner. It will also be necessary to demonstrate that autonomous AICP systems can expeditiously resolve the complexity of ethical issues that involve the safety of patients or other third parties (e.g., duty-to-warn). As I discussed previously, how AICPs derive courses of action will have increasing complexity that may pose a legal problem when needing to justify the actions of AICPs. Requirements for an audit trail with a minimum level of detail to describe decision processes are one way to help address this issue. Specifications of limits of use as well as limits of autonomy of the systems are also needed. Developers may wish to consider the need for built-in safeguards to assure that systems are only able to provide services within the boundaries of competence for the intended domain of use.

With a move toward autonomous systems, a challenge for developers of these systems is and will be how to design systems to perform ethically in diverse situations, especially when facing complex ethical dilemmas. Ethics codes are broad and general and as noted by Pope and Vasquez [41] formal ethical principles can never be substituted for an active, deliberative, and creative approach to meeting ethical responsibilities. In other words, ethics codes cannot be applied in a rote manner to autonomous artificial agents because patient's situations are unique and may call for different solutions. While intelligent machines can be designed to be highly predictable, accurate, and reliable when approaching many situations, scripted decision paths may not be sufficient in situations that entail complex ethical decision making. Thus, for intelligent machines to ever fully function unsupervised and in place of humans, these systems will need to comply with the ethical codes of conduct just as human healthcare providers currently do but also be capable of selecting the best course of action that would be considered to be reasonable based on current standards. AICP systems will also need to be able to apply theoretical approaches and expert knowledge that is current and evidence-supported.

Testing and evaluation of AICPs before they are ready for market should be a requirement. The same phases of clinical research that are used for evaluating medical devices may be one applicable model. This testing may also include a requirement to demonstrate the ethical decision making of AICPs. Ethical Turing Tests have been proposed as one way to test the ethical and moral behavior of artificial intelligent agents [42]. For example, Allen, Varner and Zinser [42] proposed the "comparative moral Turing test" (cMTT) that entails comparing the ethical decision making of an artificial intelligent agent to a person. One way to implement such a test for the mental health professions would be to compare the decisions made by an AICP to the opinions of a panel of professional mental healthcare providers.

## 6. Discussion

While there are many parallels to the ethical issues involving other modern technologies, the use of AICPs presents new and unique issues that will have significant ramifications for the mental health care and other helping professions. Additional provisions in existing ethical codes and guidelines are one necessary step to address many of the ethical issues discussed in this paper. AICP systems that are currently being developed and tested can be expected to do many of the functions that human care providers do. The use of these systems to provide information, coaching, and recommendations for "self-help" purposes may be the least restrictive use. However, more intensive treatments and therapeutic interactions will require the same ethical considerations and requirements as those between human health care providers and patients. At this

point in time, the most appropriate use of AICPs for clinical services should involve supervision by human practitioners who demonstrate competence in the use of these systems. A process whereby care providers first meet with patients and then "prescribe" use of an AICP is another model that may make sense for some clinical situations. Careful testing and evaluation of these systems will be needed regardless of application.

How and to what extent AICPs should be put to use must involve the weighing of the risks and benefits. In regards to the benefits, AICPs have the potential to greatly improve health outcomes among care seekers by customizing their care. By using machine learning techniques, AICP systems can learn the knowledge and skill sets of diverse therapeutic approaches and then deliver the most appropriate one, or adjust to different approaches based on the diagnostic profile, patient preferences, or treatment progress of patients. AICPs may also be accessed on mobile devices such as smartphones and thus be available to end users anywhere and at any time. By integrating data from other intelligent devices, such as environmental sensors and biofeedback devices, systems could further customize services to the clinical needs of patients. AICPs may also be capable of adapting to the specific aspects of the culture of the patient. For example, a virtual human psychotherapist could change its' mannerisms (e.g., eye contact), dialect of voice, use of colloquialisms, and other characteristics to match any given cultural group. This could help these systems develop and enhance rapport with patients and improve overall communication.

AICPs may also be more reliable than their human equivalents. Machines are not prone to cognitive errors, fatigue, or boredom, and they are not susceptible to forgetting things as humans do. This may result in more accurate diagnostics and attentiveness during interactions with patients. Moreover, AICPs will not succumb to care provider "burnout". Healthcare workers or caregivers who frequently interact with trauma victims, for example, can become overwhelmed and distressed themselves [43]. AICPs would have the ultimate resilience to these natural responses experienced by human caregivers. Furthermore, these systems can be immune to the personal biases that human therapists may have and thus their consultations may result in a more detached, objective point of view. Some patients may prefer encounters with an AICP rather than a human care provider for this reason, and care seekers may experience less anxiety when discussing intimate, private issues with a simulation than they would with a human care provider. This notion is supported by research that has shown preferences for the use of online counseling services among adolescents [44]. In addition, high levels of patient satisfaction with avatar-based virtual reality simulations for autism spectrum disorder skills training [45] and with the use of virtual intelligent agents in place of doctors or nurses to provide hospital discharge information [11] have been reported.

The use of machines in place of humans may in fact eliminate the risk of many of the ethical pitfalls encountered by human care providers. For example, McCorduck [46] argues that artificially intelligent judges should be used because they would be impartial and not have self-serving personal agendas. AICP systems will not have personal problems that interfere with their competence to perform work-related duties, nor will these systems have dual relationships outside of the therapeutic relationship or worse, sexual intimacies with patients or their family members.

The use of AICPs has the potential to provide economic benefits for both providers and consumers of mental health services. While potentially threatening job loss among human mental health professionals [3], use in assisting with routine tasks (including prescreening patients, etc.) can increase the efficiency of clinical practice and could potentially eliminate the need for human

mental health technicians or other support staff. AICP systems may also afford many of the same benefits as telemedicine services. For example, these systems may improve access to care services when human care providers are not available in the local area and also provide 24/7 services via smart mobile devices or personal computers with Internet connections. This would also provide cost benefits to consumers of care including not having to travel to a clinic or to attend more costly office visits. Furthermore, because AICPs can be reproduced and propagated across health care systems, they will provide significant economy of scale benefits that may provide significant cost benefits to society. If these systems help improve public health, the reduction of long-term costs to society of untreated mental health conditions as well as improved productivity resulting from a healthier population may be quite significant.

While AICPs systems have the potential to be designed and used in an ethical manner that serves the betterment of public health and wellbeing, there are also potential drawbacks to consider that extend beyond the risks already discussed in this paper. In his 1976 book *Computer Power and Human Reason* [1], Weizenbaum began to espouse concern about the use of computer technology in place of humans in professions that involve interpersonal respect, caring, understanding, and important decision making because computers lack compassion and wisdom. Weizenbaum was primarily concerned about everything in society moving toward the computational metaphor and thus imposing a mechanistic expectation on decision making that would devalue the human experience and ultimately be an affront to human dignity. Weizenbaum went so far as to state that use of computers in place of human therapists would be “immoral” [1].

Mental health care is inherently human-centered and as I noted previously, empathetic understanding between people is critical for establishing the therapeutic relationship and effective therapeutic outcomes. Patients want to feel that the psychotherapist understands them and their experiences and this can be assumed because psychotherapists are also human (thus share the same types of general feelings and experiences). It can be argued that as long as the patient is willing to interact with an artificial intelligent agent, the simulation behaves in an attentive and caring way, and the patient experiences an emotional connection with it, then that will be sufficient. However, there is clearly a difference between the espoused empathy from an artificial intelligent agent than that from another human being. The common sharing of experiences between people, and knowing that another person can experience (and probably has experienced) the same emotions and life experiences is a critical component of the interpersonal relationship. Moreover, knowing that someone else is thinking about you and cares about you can have significant therapeutic benefits. Take, for example, the evidence that brief caring messages sent from care providers that express care and empathy may potentially prevent patients from dying by suicide [47]. If the messages were sent by a “caring machine”, can we expect a patient to feel a connection with the machine and believe that machine cares about them? The same point can also be made about whether an AICP will be effective at or should be used to help people with spiritual matters. While an AICP may be capable of providing end of life counseling, for example, we have to ask the question if whether we should expect patients to be agreeable to this and whether it is moral to allow machines to do something that is inherently a human matter. Clearly, these issues have relevance for AICPs in mental health care and other helping professions such as ministry.

The psychotherapeutic process typically involves mutual discovery, a collaborative breaking down of the issues at play, the sharing of inner most thoughts, feelings, desires, and mutual discussion of the therapeutic relationship and process. The human

psychotherapist is not perfect; she/he learns from the patient and the patient knows that the psychotherapist has limitations and sometimes makes mistakes because they are human. The imperfection of the psychotherapist is, in fact, essential to the psychotherapeutic process and psychotherapy with an AICP may not be optimal if the AICP is perceived as knowing everything or is “too perfect”. Thus, as I mentioned previously in regards to simulation of forgetting and recall, it may be important for developers of AICPs to consider making these systems less than perfect. However, a perception of AICPs as superior to humans in their capabilities and knowledge, such as in the case of the “super clinician” concept, may have significant ramifications for the therapeutic relationship because the process of psychotherapy involves a process of revealing and sharing of information overtime. That is, sometimes patients will withhold information until they feel comfortable in the therapeutic relationship to disclose it. It is possible that some care seekers may feel very comfortable in openly sharing information with an AICP and do so without desiring to get to know the AICP first like they may with a human care provider. However, if a patient thinks that the AICP is capable of detecting inner thoughts (e.g., by accessing other private data or through sensing technologies), this important therapeutic process of revealing and sharing may not occur.

As noted by Weizenbaum [1], a potential consequence of the widespread and systematic use of technology in place of humans may be the risk that patients will be seen as a commodity rather than as people with individual needs. Weizenbaum paralleled this notion to the relationship between pharmaceutical companies and the field of psychiatry whereby psychiatrists have become a conduit for pharmaceutical companies to market and deliver medications to patients. This issue, which is more prominent now than in the 1970s, is partly due to the medicalization of mental health conditions, an increase in the availability of pharmacological treatments, and the preference by insurance providers for short-term interventions in place of traditional psychotherapy [48,49]. Driven by economic incentives and the perception that AICP systems are superior to traditional techniques, insurance companies may one day require the use of AICPs without giving care seekers a choice to seek care from other humans. While the increased accessibility and lower costs of AICPs may provide opportunities for longer-term treatments compared to current managed care alternatives, what may result is loss of an emphasis on key aspects of the interpersonal processes of psychotherapy and patient-centered ways of addressing everyday problems that place the needs and preferences of people first.

The public discourse that Weizenbaum took part in and inspired up until his death in 2008 is certainly relevant for the helping professions today given the highly developed capabilities of artificial intelligent agent systems and their current and projected use in mental health care and other helping professions. While some people may argue that the intent of developing AICPs is not to replace human care providers, they are in fact already replacing humans in some care providing functions. Where care-seekers may once have talked to a friend or sought counseling or treatment from a professional care provider, people can now do so online with virtual humans.

## 7. Conclusion

There are significant clinical and economic incentives for advancing the development of AICP systems. The practical uses of AICPs are far reaching, as they will serve to benefit care providers and care seekers alike. While we are not yet at the time where autonomous AICPs function as full replacements for mental health-care professionals, the issues discussed in this paper are certainly relevant for where we are now and for what can be expected

given the progressive advancements in this technological area. A move to artificial intelligent systems in place of humans may have significant ramifications for the practice of mental health care.

Many of the topics I discussed in this paper are applicable to AICPs, “helpers” and “companions” in other professional and non-professional domains (e.g., entertainment). The issues associated with the relationship between care provider and care seekers, privacy, and emotional attachment to artificial intelligent agents have common ethical and moral ramifications regardless of the specific profession or context. The ethical and moral aspects regarding the use of these systems must be well thought-out today as this will help to guide the development and use of these systems in the future. The introduction of new technologies does not necessarily mean we must jettison previous ideas; existing ethical principles and standards regarding other technologies can be applied as a guiding way forward [50]. However, new technologies also bring new types of ethical and moral questions that require very careful forethought and consideration. Professional health care organizations and regulatory boards must keep an eye on technological developments and address issues before they are abused to help guide best practices. End users, researchers, and developers, regardless of professional discipline, must all remain cognizant of the ethical and moral issues involved and be active participants in discussions and decision making regarding the current and future use of these technologies.

## Acknowledgements

I wish to thank the three anonymous reviewers for their helpful comments. I am also grateful for the assistance of Jasmine Fuller, BA, Kristine A. Johnson, PhD, Yair Neuman, PhD, Laurel D. Riek, PhD, and Kaylynn What, JD who provided comments on an earlier version of this manuscript.

## References

- Weizenbaum J. *Computer power and human reason: from judgment to calculation*. San Francisco, CA: W.H. Freeman; 1976.
- Güzeldere G, Franchi S. Dialogues with colorful “personalities” of early AI. *Stanford Hum Rev* 1995;4:161–9.
- Luxton DD. Artificial intelligence in psychological practice: current and future applications and implications. *Prof Psychol: Res Pract* 2013. <http://dx.doi.org/10.1037/a0034559>.
- Anderson M, Anderson SL, editors. *Machine ethics*. New York, NY: Cambridge University Press; 2011.
- Bostrom N. In: Smit, et al., editors. *Cognitive, Emotive, and Ethical Aspects of Decision Making in Humans and in Artificial Intelligence*, vol. 2. Oxford University; 2003. p. 12–7. International Institute of Advanced Studies in Systems Research and Cybernetics.
- Hall JS. *Ethics for Machines*; 2000. Retrieved from: <http://autogeny.org/ethics.html>
- Wallach W, Allen C. *Moral machines, teaching robots right from wrong*. Oxford University Press; 2009. ISBN: 978-0-19-537404-9.
- DeAngelis T. A second life for practice? *Monitor on psychology*; 2012, March. p. 43. Retrieved from <http://www.apa.org/monitor/2012/03/avatars.aspx>
- Rizzo AA, Lange B, Buckwalter JG, Forbell E, Kim J, Sagae K, et al. An intelligent virtual human system for providing healthcare information and support. *Study Health Technol Inform* 2011;163:503–9.
- Bickmore T, Pfeifer L. Relational agents for antipsychotic medication adherence. In: *Proceedings of the CHI'08 Workshop on Technology in Mental Health*. 2008.
- Bickmore TW, Mitchell SE, Jack BW, Paasche-Orlow MK, Pfeifer LM, O'Donnell J. Response to a relational agent by hospital patients with depressive symptoms. *Interact Comput* 2011;22(4):289–98. <http://dx.doi.org/10.1016/j.intcom.2009.12.001>.
- Defense Applied Research Projects Agency. *Detection and computational analysis of psychological signals (DCAPS)*; 2013. Retrieved from: [http://www.darpa.mil/Our\\_Work/I20/Programs/Detection\\_and\\_Computational\\_Analysis\\_of\\_Psychological\\_Signals\\_\(DCAPS\).aspx](http://www.darpa.mil/Our_Work/I20/Programs/Detection_and_Computational_Analysis_of_Psychological_Signals_(DCAPS).aspx)
- InTouch Health. *RP-VITA robot*; 2012. Retrieved from: <http://www.intouchhealth.com/products-and-services/products/rp-vita-robot/>
- Ferrucci D, Levas A, Bagchi S, Gondek D, Mueller ET. *Watson: beyond jeopardy!* *Artif Intell* 2013;199:93–105.
- IBM. *IBM Watson: ushering in a new era of computing*; 2014. Retrieved from: <http://www-03.ibm.com/innovation/us/watson/index.shtml>
- Corey G, Corey MS, Callanan P. *Issues and ethics in the helping professions*. Cengage Learn 2007.
- Koocher GP, Keith-Speigel P. *Ethics in psychology and the mental health professions: standards and cases (oxford textbooks in clinical psychology)*. USA: Oxford University Press; 2008.
- American Psychological Association. *American Psychological Association ethical principles of psychologists and code of conduct*. American Psychiatric Association; 2002. Retrieved from: <http://www.apa.org/ethics/code2002.html> [accessed on 09.02.14].
- American Medical Association. *Opinion 1.02 – the relation of law and ethics*; 1994. Retrieved from: <http://www.ama-assn.org/ama/pub/physician-resources/medical-ethics/code-medical-ethics/opinion102.page>
- Veruggio G. The euron roboethics roadmap. In: *6th IEEE-RAS international conference on humanoid robots*. IEEE; 2006, December. p. 612–7.
- Veruggio G, Operto F. *Roboethics: social and ethical implications of robotics*. Springer Handbook of Robotics; 2008. p. 1499–524.
- Asimov I. *Runaround*. In: *Astounding science fiction*. Street and Smith Publications, Inc.; 1942.
- Asimov I. *Robots and empire*. New York, NY: Doubleday; 1985.
- Riek LD, Howard D. *A code of ethics for the human–robot interaction profession*. In: *Proceedings of We Robot*. 2014.
- Moor JH. The nature, importance, and difficulty of machine ethics. *IEEE Intell Syst* 2006;21(4):18–21.
- Bordin ES. The generalizability of the psychoanalytic concept of the working alliance. *Psychother: Theory Res Pract* 1979;16(3):252–60.
- Gelso CJ, Carter J. Components of the psychotherapy relationship: their interaction and unfolding during treatment. *J Counsel Psychol* 1994;41:296–306.
- Horvath AO, Symonds BD. Relation between working alliance and outcome in psychotherapy: a meta-analysis. *J Counsel Psychol* 1991;38(April (2)):139–49. <http://dx.doi.org/10.1037/0022-0167.38.2.139>.
- American Psychiatric Association. *The principles of medical ethics with annotations especially applicable to psychiatry*; 2013. Retrieved from: <http://www.psych.org/File%20Library/Practice/Ethics%20Documents/principles2013-final.pdf> [accessed on 09.02.14].
- Wolfendale J. My avatar, my self: virtual harm and attachment. *Ethics Inform Technol* 2007;9(2):111–9. <http://dx.doi.org/10.1007/s10676-006-9125-z>.
- Rogers C. *A theory of therapy, personality and interpersonal relationships as developed in the client-centered framework*. In: Koch S, editor. *Psychology: a study of a science*. Vol. 3: formulations of the person and the social context. New York: McGraw Hill; 1959.
- Whalley LJ. *Ethical Issues in the application of virtual reality to medicine*. *Comput Biol Med* 1995;25(2):107–14.
- Miller KW. It's not nice to fool humans. *IT Prof* 2010;12(1).
- Riek LD, Watson RNW. The age of avatar realism. *IEEE Robot Autom* 2010;17(4).
- Kramer GM, Mishkind MC, Luxton DD, Shore JH. Managing risk and protecting privacy in telemental health: an overview of legal, regulatory, and risk management issues. In: Myers, Turvey, editors. *Telemental health: clinical, technical and administrative foundations for evidence-based practice*. Waltham, MA: Elsevier; 2013. p. 83–107.
- Sullins JP. *When is a robot a moral agent?* In: *Machine ethics* Andersen & Andersen. New York, NY: Cambridge University Press; 2011.
- Moore A. Intangible property: privacy, power and information control. In: Moore A, editor. *Information ethics: privacy, property, and power*. Seattle, WA: University of Washington Press; 2005.
- BBC News. *US spy leaks: How intelligence is gathered*; 2014, January. Retrieved from: <http://www.bbc.com/news/world-us-canada-24717495> [accessed on 19.02.14].
- Luxton DD, O'Brien K, McCann RA, Mishkind MC. Home-based telemental healthcare safety planning: what you need to know. *Telemed e-Health* 2012;18(8):629–33.
- Richards D, Bransky K. ForgetMeNot: what and how users expect intelligent virtual agents to recall and forget personal conversational content. *Int J Hum-Comput Stud* 2014;72(5):460–76. <http://dx.doi.org/10.1016/j.ijhcs.2014.01.005>.
- Pope KS, Vasquez MJ. *Ethics in psychotherapy and counseling: a practical guide*. 4th ed. Wiley; 2010.
- Allen C, Varner G, Zinser J. Prolegomena to any future artificial moral agent. *J Exp Theor Artif Intell* 2000;12(3):251–61. <http://dx.doi.org/10.1080/09528130050111428>.
- Figley CR. *Compassion fatigue as secondary stress disorder: an overview. Compassion fatigue: coping with secondary traumatic stress disorder in those who treat the traumatized (1–20)*. New York: Brunner/Mazel; 1995.
- King R, Bambling M, Lloyd C, Gomurra R, Smith S, Reid W, et al. Online counselling: the motives and experiences of young people who choose the Internet instead of face to face or telephone counselling. *Counsel Psychother Res* 2006;6(3):169–74.
- Kandalaf MR, Didehbani N, Krawczyk DC, Allen TT, Chapman SB. Virtual reality social cognition training for young adults with high-functioning autism. *J Autism Dev Disord* 2012. <http://dx.doi.org/10.1007/s10803-012-1544-6>.
- McCorduck P. *Machines who think*. 1st ed. W. H. Freeman; 1979.

- [47] Luxton DD, June JD, Comtois KA. Can post-discharge follow-up contacts prevent suicide and suicide behavior? A review of the evidence. *Crisis: J Crisis Interv Suicide Prev* 2013;34:32–41, <http://dx.doi.org/10.1027/0227-5910/a000158>.
- [48] Clay RA. Advocating for psychotherapy, vol. 42(8). American Psychological Association; 2011, September. Retrieved from: <https://www.apa.org/monitor/2011/09/psychotherapy.aspx>
- [49] Olfson M, Marcus SC. National trends in outpatient psychotherapy. *Am J Psychiatry* 2010;167:1456–63, <http://dx.doi.org/10.1176/appi.ajp.2010.10040570>.
- [50] Grimes JM, Fleischman KR, Jaeger PT. Virtual guinea pigs: ethical implications of human subjects research in virtual worlds. *Int J Internet Res Ethics* 1999;2(1):38–56.