Artificial Intelligence in Medicine 62 (2014) 1–10

Contents lists available at ScienceDirect

Artificial Intelligence in Medicine

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

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

David D. Luxton[a,b,*]

[a] Department of Psychiatry and Behavioral Sciences, University of Washington School of Medicine, Box 356560, Seattle, WA 98195, United States
[b] 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

http://dx.doi.org/10.1016/j.artmed.2014.06.004
0933-3657/Published by Elsevier B.V.

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), 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