Ethical reasoning in technology: using computational approaches to integrate ethics into AI systems

Sahil Sholla and Iraq Ahmad Reshi
Department of CSE, Islamic University of Science and Technology, Awantipora, India

Abstract

Purpose – This paper does not concern with the "why" of ethics. Such questions are typically of interest to philosophers and are outside the scope of this work. In the next section, the authors offer a look into "what" of ethics, i.e. various types and subtypes of ethics. Subsequently, the authors explore "how" of ethics, by summarising various computational approaches to ethical reasoning offered by researchers in the field.

Design/methodology/approach – The approaches are classified based on the application domain, ethical theory, agent type and design paradigm adopted. Moreover, promising research directions towards ethical reasoning are also presented.

Findings – Since the field is essentially interdisciplinary in nature, collaborative research from such areas as neuroscience, psychology, artificial intelligence, law and social sciences is necessary. It is hoped that this paper offers much needed insight into computational approaches for ethical reasoning paving way for researchers to further engage with the question.

Originality/value – In this paper, the authors discussed vaious computational approaches proposed by researchers to implement ethics. Although none of the approaches adequately answer the question, it is necessary to engage with the research effort to make a substantial contribution to the emerging research area. Though some effort has been made in the design of logic-based systems, they are largely in stages of infancy and merit considerable research.

Keywords Artificial intelligence, Normative ethics, AI ethics, Fuzzy variables, Utilitarian ethics

Paper type Literature review

1. Introduction

Machines capable of exhibiting moral ethical behaviour have traditionally been a subject of interest for philosophers and science fiction writers. However, as artificial intelligence (AI)-enabled technologies increasing integrate in our daily lives, the question of designing machines capable of ethical reasoning becomes inescapable. Until recently, the relationship between computing and ethics was only concerned with the responsible use of technology, i.e. hacking, software property rights, privacy issues, etc. However, increased presence of smart devices in our daily lives necessitates teaching manners to machines. This is particularly a concern as AI technologies advance to enable autonomous systems capable of executing decisions with ethical ramifications.

The ever-increasing presence of modern technologies in our lives inadvertently shapes our social, moral and ethical milieu (Jasanoff, 2016; Machidon, 2015). Although advanced technological developments are expected to extend conveniences and comfort, concerns about unethical behaviour of intelligent machines also loom (O'Neil, 2016). Many instances

---

Conflict of interest: The authors declare that they have no conflict of interest.