ხელოვნური მორალური აგენტი

სამაგისტრო ნაშრომი

ავტორი: ლაშა უსუფაშვილი

ხელმძღვანელი: თამარ ცხადაძე

მეცნიერების და ხელოვნების ფაკულტეტი

ილიას სახელმწიფო უნივერსიტეტი

26.06.2019

Abstract

In our time technologies penetrates every sphere of human activities. The moral sphere is no different. It's become increasingly important to analyze whether it is possible to have artificially made moral agent or not. Reason for that is increasing in usage of technological artifacts which makes it impossible for them, to be excluded from moral sphere. To answer this question there are two distinct ways to do so. First explores possibility of different moral agent other than human. Second is to model artificial agent using the AI computer system. Both have its pros and cons, but at the end both failed to accomplish impossible. The impossible is to give machine intentions without it, it is impossible to have an agent which acts in accordance of its own will, which is vital for morality, as described by Campbell. Only viable option, then becomes to simulate moral deliberation and knowing that it is a limited one. I call it an antipodean moral agent. Agent without self-consciousness and own intentions, an agent which mimics to be conscious and moral without any understanding of variables it manipulates.

Keywords:

Artificial moral agent, Modeled moral deliberation, Chinese room argument, Consciousness in artificial agent, Artificial agent, morality, Moral deliberation.

აბსტრაქტი

ჩვენ ოპოქაში ტექნოლოგიების გამოყენების არეალი იზრდება ყოველწამიერად. მიკრო-ჩიპებით თუ ნულებისა და ერთების კომბინაციით მართულ სამყაროში, ბუნებრივია ტექნოლოგიები მორალის სფეროშიც შემოდიან. ამ ფონზე მნიშვნელოვანი ხდება ხელოვნური მორალური აგენტობის დადგენა. ამ მიზნით არსებობს ორი ძირითადი მიდგომა. ერთი სწავლობს ზოგადად ნებისმიერი სახის, ადამიანისგან განსხვავებულ მორალურ აგენტობას, ხოლო მეორე კი ადამიანის მსგვასი, მორალური აგენტის ხელოვნურად შექმნის შესაძლებლობას. ორივე მიმართულებას აქვს თავის უპირატესობები, ისევე როგორც ნაკლოვანებები. იმისათვის, რომ დავადგინოთ თუ რომელ მიმართულებას შეუძლია ამ თეორიული საკითხის გადაჭრა, რაც ნაშრომის მთავარი მიზანია, საჭიროა დეტალურად მიმოვიხილოთ ისინი. მიდგომები, რომლებიც მიმართულია მოახდინონ ადამიანის მორალური განსჯის მოდელირება, ვერ ახარხებენ ამას, რადგან მათ არ შეუძლიათ მორალურობის მთავარი ასპექტის, განზრახულობის კომპიუტერულ სისეტმაში შექმნა. მეორეს მხრივ, სხვა ტიპის აგენტობის არცერთი თეორია არ არსებობს, რომელსაც შეუძლია დაასაბუთონ, უკანსკნელის შესაძლებლობა. შედეგად ჩვენ ვრჩებით ისეთი ხელოვნური აგენტობის იმედად, რომელიც არ არის ნამდვილი და შეიძლება მხოლოდ სიმულაციად ჩაითვალოს.

საკვანძო სიტყვები:

ხელოვნური მორალური აგენტი, მორალური განსჯის მოდელირება, სხვა ტიპის მორალური აგენტი, ხელოვნური ინტელექტი, ცნობიერება, ჩინური ოთახის არგუმენტი.