| | |
|---|---|
| 氏　　名 | NIEUWAZNY JAGNA STANISLAWA |
| 授与学位 | 博士（工学） |
| 学位記番号 | 博甲第189号 |
| 学位授与年月日 | 令和3年3月19日 |
| 学位授与の要件 | 学位規則第4条第1項 |
| 学位論文題目 | A Study on Implementing of Culture, Religion and Time-Awareness to Machine Ethics Algorithms<br>（文化・宗教・時間経過認識の機械倫理アルゴリズムへの実装に関する研究） |
| 論文審査委員 | 主査　教授　桝井　文人<br>准教授　プタシンスキ　ミハウ　エドムンド<br>教授　升井　洋志<br>教授　阿部　良夫<br>准教授　渡辺　美知子 |

# 学位論文内容の要旨


Recent rapid developments in the field of Artificial Intelligence (AI), especially those regarding the development and implementation of artificial neural networks, surface the questions previously considered to arise much later in the future and were so far only addressed in philosophy, or its lighter derivatives, such as science fiction literature. One of such questions regards the limitations of Artificial Intelligence systems to fully grasp the importance and necessity in human lives of such subjective, spiritual and motivational phenomena as religious experience or a moral outlook. In this dissertation, I present the results of my research devoted to applying Artificial Intelligence to quantitative analysis of factors influencing the ethical outlook of Japanese people.

I begin with presenting the background of this research.
In particular, I focus on the relations between Artificial Intelligence and ethics- seen as a set of moral rules transmitted through religion, education or historical experience.

After that, I report the results of experiments conducted in the course of my research. As an introduction to the first experiment, I provide an overview of previous cross-sectional research in the field of Religion and Technology. I then analyze how much religious vocabulary, in particular Buddhist vocabulary taken from the largest online dictionary of Buddhist terms, is present in everyday social space of Japanese people, particularly, in Japanese blog entries appearing on a popular blog service (Ameba blogs).

I interpreted the level of everyday usage of Buddhist terms as appearance of such terms in the consciousness of people. I further analyzed what emotional and moral associations such contents generate. In particular, I analyzed whether expressions containing Buddhist vocabulary are considered appropriate or not from a moral point of view, as well as the emotional response of Internet users to Buddhist terminology.

Secondly, I focus on ethical education as a means to improve artificial companion's conceptualization of moral decision-making process in human users. In particular, I focus on automatically determining whether changes in ethical education influenced core moral values in humans throughout the century. I analyze ethics as taught in Japan before WWII and today to verify how much the pre-WWII moral attitudes have in common with those of contemporary Japanese, to what degree what is taught as ethics in school overlaps with the general population's understanding of ethics, as well as to verify whether a major reform of the guidelines for teaching the school subject of "ethics" at school after 1946 has changed the way common people approach core moral questions (such as those concerning the sacredness of human life). I selected textbooks used in teaching ethics at school from between 1935 and 1937, and those used in junior high schools today (2019) and analyzed what emotional and moral associations such contents generated. The analysis was performed with an automatic moral and emotional reasoning agent and based on the largest available text corpus in Japanese as well as on the resources of a Japanese digital library.

Finally, in the course of a third experiment and based on the findings of the two previous ones, I try to answer a twofold question.

Firstly, since the methods used for performing authorship analysis imply th at an author can be recognized by the content he or she creates, I was intere sted in finding out whether it would be possible for an author identification solution to correctly attribute works to authors if in the course of years they have undergone a major psychological transition.

Secondly - and from the point of view of the evolution of an author's ethical values - I checked what it would mean if the authorship attribution system encounters difficulties in detecting single authorship, hypothesizing that it could mean that historical events had had significant impact on the person's ethical outlook. I set out to answer those questions through performing a binary authorship analysis task using a text classifier based on a pre-traine d transformer model.

Based on the findings from those experiments, I outline future research goals.

# 論文審査結果の要旨

人工知能に関連する急速な技術発展により，最近まで対象外と考えられてきた様々なタスクが議論の対象となりつつある。宗教観や道徳観，倫理観を計算機で扱う，あるいは情報科学的アプローチによりそれらを分析するといったタスクもそのひとつと言える。著者は，大規模コーパス，倫理推論エージェント，機械学習モデルを活用することで、日本における宗教や教育，歴史的経験がある時代の道徳的価値観に与える影響や相互の関連性を自然言語処理および機械学習モデルという計算機科学的アプローチを用いて定量的に観察することにより，仏教に由来する日本語の特性や語義変化を明らかにするとともに，集団としての日本人が共有する道徳観／倫理観に対して歴史的変化が与える影響，個人の道徳観／倫理観に対して歴史的事象が与える影響などを明らかにした。

これを要するに，著者は，計算機科学において昨今課題となりつつある機械倫理に関連して，機械倫理アルゴリズム化の可能性について示唆を与える新知見を得たものである。自然言語処理，人工知能，認知科学など複数の分野に跨る深淵な課題に対して貢献するところ大なるものがある。よって著者は、北見工業大学博士（工学）の学位を授与される資格があるものと認める。