様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　３ 年　 ５ 月 ２７ 日現在

機関番号：１２６０１
研究種目：挑戦的研究（萌芽）
研究期間：2018～2020
課題番号：１８Ｋ１８４３４

研究課題名（和文）人工知能の内部仕様に結合された人工知能倫理の動的構成

研究課題名（英文）Organizing AI ethics dynamically with the connection to the internal AI specifications

研究代表者
堀　浩一（Hori, Koichi）

東京大学・大学院工学系研究科（工学部）・教授

研究者番号：４０１７３６１１

交付決定額（研究期間全体）：（直接経費）　　4,400,000円

研究成果の概要（和文）：近年の人工知能の研究開発の進展が社会に及ぼす影響は甚大であると予想されており、人工知能をめぐるELSI(Ethical Legal Social Issues)に関する議論が活発に行われるようになり、国を超えた協力も始まりつつある。しかしながら、それらの人工知能倫理に関わる問題の議論に参加している人工知能研究の専門家はごく少数に限られており、倫理的な問題の議論が人工知能の研究開発に十分に生かされているとは言い難い。本研究においては、人工知能倫理の問題と人工知能の仕様との間の橋渡しをするシステムを構築した。実験の結果、倫理的問題を考慮すると、かえって創造的設計が生まれるという効果が確認された。

研究成果の学術的意義や社会的意義
　学術的には、複雑化するAI倫理の問題を系統的にとらえることを我々が作成したシステムが可能にした。また、人工知能技術を人工知能の抱える問題に自己適用することにも成功した。
　社会的意義としては、倫理的な問題の議論と人工知能の設計開発を、我々の提案したシステムにより、結合することが可能となった。これまで倫理的な問題は自由な研究開発を抑圧するという誤解もしばしば見受けられたが、我々の研究によれば、倫理的問題を考慮すると、むしろこれまでよりも創造的な設計解が得られることが明らかになった。

研究成果の概要（英文）：Recent development of AI research has caused concerns about ethical issues. Much discussion is in progress all over the world. However, those discussions are not well reflected to the design of AI systems.This study proposes a system which bridges the gap between AI specifications and AI ethics.
We have proved that AI researchers and developers can design AI systems more creatively when they properly consider ethical issues supported by our system.

研究分野：人工知能

キーワード：人工知能倫理　創造活動支援システム　設計支援

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９（共通）

１．研究開始当初の背景

近年の人工知能の研究開発の急速な進展が社会に及ぼす影響は甚大であると予想されており、日米欧のそれぞれにおいて、人工知能をめぐる ELSI(Ethical Legal Social Issues)に関する議論が活発に行われるようになり、国を超えた協力も始まりつつある。（以下、本報告においては、記述を簡潔にするために、それらの議論で扱われている問題を代表するキーワードとして、「人工知能倫理」というキーワードを用いることにする。扱うべき問題の範囲は広いが、たとえば、人工知能学会においてそれらの広い問題を議論している委員会の名称も、倫理委員会であるので、倫理というキーワードにそれらの広い問題を代表させることにしたい。）

しかしながら、それらの人工知能倫理に関わる問題の議論に参加している人工知能研究の専門家はごく少数に限られている。その理由は、多くの場合、自分の抱えている人工知能の研究開発の仕事が忙しくて、倫理の問題に関わっていられない、というものであろうと想像されるが、倫理に関わる議論が「自由な研究開発を抑圧するものである」と誤解している研究者や開発者も少なくないのが実情である。

多くの人工知能の研究者あるいは開発者は、世界的な人工知能倫理に関する議論を、自分が研究開発している人工知能にどう生かせばよいのか、理解できていない。たとえば、「人工知能は透明性を有することが望ましい」というガイドラインが人工知能倫理に関連して提示された時に、人工知能研究開発者の多くは、自分の構築する人工知能の何をどのように変更すれば透明性の要求に応えたと言えるのかを理解できない。すなわち、人工知能の内部仕様の記述と人工知能倫理の記述の間に大きな乖離が存在し、その乖離を埋めるための方策が何も与えられていないのが現在の状況である。また、人工知能倫理をめぐる議論が世界的に膨大になりつつあり、現場の仕事に忙しい人工知能研究開発者がそれらの全体像と詳細な相互関係を把握できなくなっているという問題も存在している。

２．研究の目的

本研究の目的は、人工知能の内部仕様の記述と人工知能倫理の記述の間の乖離を埋めるための方 策の一つを与えることである。技術哲学における社会構成主義として知られているように、技術が 社会を決めるのでもなく、社会が技術を決めるのでもなく、技術と社会は相互に作用する。しかし、 残念ながら、現在は、人工知能の技術の進歩のスピードと人工知能倫理の議論のスピードとの違い が大きいために、人工知能倫理の議論を生かさないままに人工知能の技術の研究開発がどんどん進みつつある。本研究では、人工知能の内部仕様の設計と人工知能が社会に及ぼす影響に関わる人工 知能倫理の議論との間の橋渡しをするプラットフォームを構築することを目指す。

人工知能を要素に含む新しい人間社会の構成に関わる人工知能倫理の議論も、人工知能を構成す るための内部仕様に直接関わる技術も、どちらも変化しつづけている。したがって、本研究で構築 するプラットフォームも、それらの動的変化に対応し、さらには好ましい動的変化を促進するようなものでなければならない。扱うべき情報の量も膨大になりつつある現在、一人の人間でそれらを全て把握することは困難になりつつある。本研究においては、その困難さを解決するためにもまた 人工知能の技術が役に立つと考え、人工知能の技術を活用して、上に述べたプラットフォームを構築する。

３．研究の方法

人工知能倫理と人工知能の内部仕様の設計の間の橋渡しをするプラットフォームを構築するための手順は次のとおりである。

1．人工知能倫理に関わる議論のドキュメントを継続的に収集する。

2．人工知能の内部仕様を構成する人工知能技術のドキュメントを継続的に収集する。

3．人工知能倫理に関わるドキュメント群に対してテキストマイニングを行い、人工知能倫理に関 する概念が構成する概念空間を動的に構成する。

4．人工知能技術に関わるドキュメント群に対してテキストマイニングを行い、人工知能技術に関わる概念が構成する概念空間を動的に構成する。

5．人工知能倫理の概念空間と人工知能技術の概念空間を結合するための橋渡しの概念空間を、筆者らが研究してきたトピックブリッジングの手法を用いて動的に構成する。

6．人工知能倫理の概念空間、人工知能技術の概念空間、および橋渡しの概念空間を利用しながら 人工知能の外部仕様と内部仕様を考えるための設計支援の機能を提供する。この設計支援の機能の実現のためには、筆者らが研究してきた創造的設計支援システムの研究成果を応用する。

４．研究成果

上に述べた通りの方法で、システムを構築した。

人工知能の設計者が、このシステムに、自分の作ろうとする人工知能システムの仕様を自然言語の文章として入力すると、システムは、まずその仕様がどういう概念構造を持っているかをグラフィックに表示する。それを図１に示す。

次に、その人工知能システムが社会にどのような影響を及ぼしうるかの可能シナリオをシステムが自動的に提案する。さらに、その上位のシナリオを活かして AI の仕様をシステムと対話しながら考え直すことができる。それを図２に示す。

たとえば、高度な画像認識を応用したシステムを開発しようとしていた設計者が、そのシステムの概要を自然言語の文章の形でこのシステムに入力すると、システムからは、その画像認識が個人のプライバシーを侵害することにつながるかもしれないというような可能シナリオなどが提示される。ユーザは、プライバシーの問題のさらに上位にある概念なども確認しながら、自分が作ろうとしていたシステムの価値を、倫理レベルから考え直すことになる。

倫理レベルから考え直すと、技術シーズにドライブされて新たな AI システムの仕様を考えていた技術者は、人々への貢献の価値のレベルで新しいアイディアに思い至り、技術的な仕様そのものにも、新しい解を考え始める。従来も、優秀な技術者は自分でそのループを回していたのであるが、我々のシステムは最新の倫理的問題の知識ベースを有しているので、我々のシステムを用いれば、ユーザは、的確に思考空間を広げることができる。

この実験システムの効果を確認するためのユーザスタディを実施し、その効果の分析を行った。

プロの研究者、開発者、ベンチャー企業経営者などに、このシステムを使ってもらった。

その結果、当初の期待以上の成果を得ることができた。まず、当初目指していた通りに、倫理的な問題に興味のなかった人工知能研究者たちに、自分の研究や開発においてどのような倫理的な問題が潜んでいるのかを、考えてもらうことができるようになった。しかも、それだけにとどまらず、倫理的な問題に思いを致すと、それまでよりも、創造的な設計解が得られる、という嬉しい効果が得られた。これは、倫理的な問題という制約が加わることにより、設計者が新しい設計解空間を探し始めることによる。

それらの成果を２編の国際学術誌のオープンアクセス論文や国際会議論文などとして発表した。それらの論文は国際的にも注目を集め、一時期、Altmetrics で全世界の全分野の論文でトップ10%の論文という評価を得た。

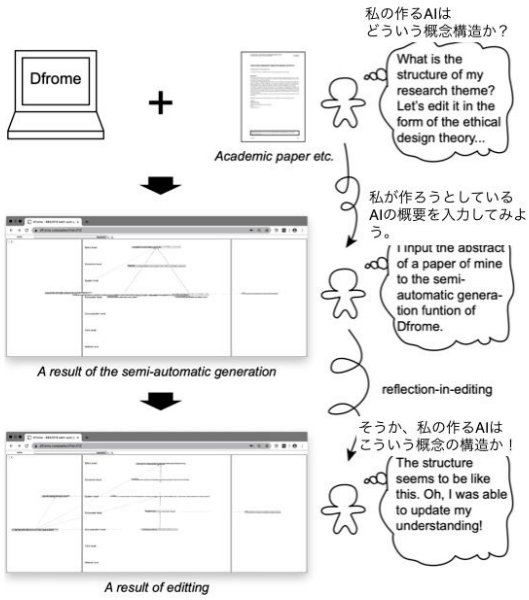

図１　ユーザが自分の作ろうとする AI システムの仕様を自然言語の文章として入力すると、システムは、その概念構造をグラフィック表示する。

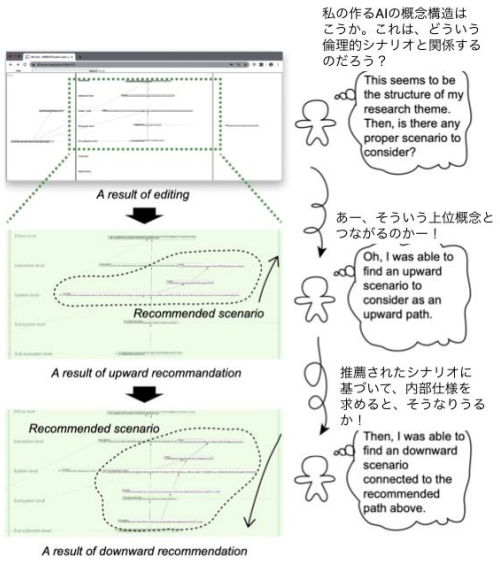

図２ ユーザの入力した AI システムの仕様が、どういう倫理レベルの問題につながりうるかの可能シナリオをシステムが自動的に提案する。また、そのシナリオに沿って AI システムの仕様を考え直すことができる。

５．主な発表論文等

〔雑誌論文〕　計3件（うち査読付論文　2件／うち国際共著　0件／うちオープンアクセス　3件）

１．著者名 Kaira Sekiguchi and Koichi Hori	４．巻 36
２．論文標題 Designing ethical artifacts has resulted in creative design.	５．発行年 2021年
３．雑誌名 AI and Society	６．最初と最後の頁 101-148
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/s00146-020-01043-6	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

１．著者名 Sekiguchi Kaira、Hori Koichi	４．巻 online20180ct
２．論文標題 Organic and dynamic tool for use with knowledge base of AI ethics for promoting engineers ' practice of ethical AI design	５．発行年 2018年
３．雑誌名 AI & SOCIETY	６．最初と最後の頁 1-21
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/s00146-018-0867-z	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

１．著者名 堀　浩一	４．巻 2
２．論文標題 人工知能として認識されない人工知能の埋め込まれる社会に向けて	５．発行年 2018年
３．雑誌名 情報通信政策研究	６．最初と最後の頁 11-19
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

〔学会発表〕　計5件（うち招待講演　1件／うち国際学会　4件）

１．発表者名 Kaira Sekiguchi and Koichi Hori
２．発表標題 Networking AI systems to promote ethical design practice
３．学会等名 he second edition of the international Philosophy of Human-Technology Relations confer- 161 162 論文目録 ence, PHTR2020（国際学会）
４．発表年 2020年

１．発表者名
Kaira Sekiguchi and Koichi Hori

２．発表標題
`Can ethics enhance creative design activity?

３．学会等名
ICED19 22nd International Conference on Engineering　Desig（国際学会）

４．発表年
2019年

１．発表者名
Kaira Sekiguchi and Koichi Hori

２．発表標題
Realization of Organic and Dynamic Creativity Support Tool for Promoting Ethical AI Design

３．学会等名
13th International Conference on Knowledge, Information and Creativity Support Systems (KICSS2018)（国際学会）

４．発表年
2018年

１．発表者名
Koichi Hori

２．発表標題
Creativity Support <-> AI Ethics

３．学会等名
UNESCO Roundtable of `Artificial Intelligence: Reflection on its complexity and impact on our society'（招待講演）

４．発表年
2018年

１．発表者名
Jason Millar, Brent Barron, Koichi Hori, Finlay, Kentaro Kotsuki, Ian Kerr

２．発表標題
Accountability in AI - Promoting Greater Social Trust

３．学会等名
G7 Multi-stakeholder Conference on Artificial Intelligence: Enabling the Responsible Adoption of AI（国際学会）

４．発表年
2018年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

―

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考

７．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

８．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関