

戦略的創造研究推進事業  
(社会技術研究開発)  
令和2年度研究開発実施報告書

「人と情報のエコシステム」

研究開発領域

「人と情報テクノロジーの共生のための  
人工知能の哲学2.0の構築」

鈴木 貴之

(東京大学大学院総合文化研究科 准教授)

## 目次

|                                      |    |
|--------------------------------------|----|
| 1. 研究開発プロジェクト名 .....                 | 2  |
| 2. 研究開発実施の具体的内容 .....                | 2  |
| 2 - 1. 研究開発目標 .....                  | 2  |
| 2 - 2. 実施内容・結果 .....                 | 2  |
| 2 - 3. 会議等の活動 .....                  | 6  |
| 3. 研究開発成果の活用・展開に向けた状況 .....          | 7  |
| 4. 研究開発実施体制 .....                    | 7  |
| 5. 研究開発実施者 .....                     | 8  |
| 6. 研究開発成果の発表・発信状況、アウトリーチ活動など .....   | 9  |
| 6 - 1. シンポジウム等 .....                 | 9  |
| 6 - 2. 社会に向けた情報発信状況、アウトリーチ活動など ..... | 9  |
| 6 - 3. 論文発表 .....                    | 10 |
| 6 - 4. 口頭発表（国際学会発表及び主要な国内学会発表） ..... | 10 |
| 6 - 5. 新聞／TV報道・投稿、受賞等 .....          | 11 |
| 6 - 6. 知財出願 .....                    | 11 |

## 1. 研究開発プロジェクト名

人と情報テクノロジーの共生のための人工知能の哲学2.0の構築

## 2. 研究開発実施の具体的内容

### 2 - 1. 研究開発目標

- ①人工知能の可能性と限界を考察するうえで鍵となる徳を中心とした諸概念の関係を明らかにし、哲学研究者以外にもその成果が容易に理解できるように、コンセプト・マップなどを作成する。
- ②人工知能の社会実装可能性を考えるための手がかりとなる概念枠組を構築する。具体的には、人間と人工知能のとりうる関係を類型化し、各類型の実現可能性や長所・短所などを明らかにしたチャートなどを作成する。
- ③人工知能の可能性と限界を検討するための新たな理論的枠組（人工知能の哲学2.0）を構築する。その内容を、人工知能研究者や人工知能の社会実装に携わる人々もアクセスできる教科書や概説書などの形で公刊する。
- ④情報テクノロジー研究開発者へのインタビューや研究会の開催などを通じて、情報テクノロジーの研究開発において、哲学をはじめとする人文諸科学に（倫理的問題の検討以外に）どのような貢献の可能性があるかを明らかにする。
- ⑤学会ワークショップやシンポジウムの開催などを通じて、情報テクノロジーの研究開発者と人文科学研究者との交流を促進する。さらに、一般向けのトークイベントの開催や新聞・雑誌における記事の執筆などを通じて、研究者や技術者と一般の人々との問題関心の共有可能性を探る。

### 2 - 2. 実施内容・結果

#### (1) スケジュール

| 実施項目                    | 2018年度 | 2019年度 | 2020年度 | 2021年度 |
|-------------------------|--------|--------|--------|--------|
| 1. 人工知能研究の歴史と現状のレビュー    | →      | →      | →      |        |
| 2. 人工知能の哲学の体系的再検討       | →      | →      |        |        |
| 3. 徳概念を手がかりとした各論点の関係の分析 |        | →      | →      |        |
| 4. 人工知能の可能性と限界の解明       |        |        | →      |        |
| 5. 成果物の作成               |        |        |        | →      |

|                              |   |   |   |   |
|------------------------------|---|---|---|---|
| 6. 徳に関する哲学的議論のレビュー           | → | → |   |   |
| 7. 経験的知見をふまえた徳理解のアップデート      |   | → |   |   |
| 8. 人工知能による徳の実現可能性            |   |   | → |   |
| 9. 成果物の作成                    |   |   |   | → |
| 10. 拡張された心に関する先行研究のレビュー      | → | → |   |   |
| 11. 情報テクノロジーによる人間知性の拡張可能性の検討 |   | → | → |   |
| 12. 人間知性と人工知能の協働可能性の検討       |   |   | → |   |
| 13. 成果物の作成                   |   |   |   | → |

## (2) 各実施内容

### 今年度の到達点①：

(目標) 人工知能研究の現状に関する理解を深める

実施項目1：人工知能研究の歴史と現状のレビュー（全体計画書の実施項目1）

実施内容：①プロジェクトメンバーおよび東京大学の教員・大学院生をメンバーとして、深層学習および人工知能研究全般に関する読書会を実施し、人工知能研究の現状に関する理解を深めた。②2021年3月に、三宅陽一郎氏および中島秀之氏へのインタビューを実施した。

### 今年度の到達点②：

(目標) 現在の人工知能の可能性と限界の哲学的分析

実施項目2-1：人工知能の可能性と限界の解明（全体計画書の実施項目4）

実施内容：①上記実施項目1で得た知見をふまえて、現在の人工知能の可能性と課題を考察し、その成果を鈴木貴之が『科学哲学』掲載論文「深層学習の哲学的意義」として公刊した。②人工知能の研究開発がもつ2つの可能性について検討し、2020年11月にオランダのトゥウェンテ大学が開催した国際学会 Philosophy of Human-Technology Relations Conference 2020 (PHTR2020) におけるパネルセッション "Artificial Intelligence as a Tool" において、鈴木貴之が "Two Conceptions of Artificial Intelligence" として発表を行った。

実施項目2-2：人工知能による徳の実現可能性（全体計画書の実施項目8）

実施内容：①立花幸司が、倫理的な徳を構成する各要素のどのような側面が人工知能によって涵養可能であるのか、またどのような側面は実現が困難であるかを明らかにし、その成果を上記PHTR2020において発表した。②植原亮が、認知的徳の場合における同様の論点を単著や論文を通じて明らかにした。③これらの成果を元にしつつ、染谷グループの中澤栄輔と連携し、より一般向けの情報発信としてワークショップを開催し、人工知能による徳の実現可能性について広く議論をおこなった。

実施項目2-3：情報テクノロジーによる人間知性の拡張可能性の検討（全体計画書の実施項目11）

実施内容：①2020年8月に、一橋大学の久保明教氏を講演者とした知性の拡張に関するオンライン講演会を開催した。②技術哲学およびテクノロジーの人類学における関連文献の調査を実施した。

実施項目2-4：人間知性と人工知能の協働可能性の検討（全体計画書の実施項目12）

実施内容：①技術哲学における関連文献の調査と分析を引き続き実施した。②上記PHTR2020のパネルセッションにおいて、柴田崇が、AI (artificial intelligence) とIA (intelligence amplifier) という2つの構想を比較し、"AI vs. AI: The Real Issues Hidden in the Struggle"として発表した。また、上杉繁が、道具の2つのタイプという観点から人工知能の可能性を検討し、"Considerations on Analysing Relations between Humans and AI Technologies Based on Archetypes of Instruments – Club-type and Pot-type"として発表した。

実施項目2-5：ウェルビーイングに関する哲学的検討（全体計画書未記載項目）

実施内容：人間と情報テクノロジーの共生可能性を考えるための理論的基礎の確立およびERATO池谷プロジェクトとの連携準備のために、ウェルビーイングに関する哲学的検討を行った。具体的には、①鈴木貴之、植原亮、染谷昌義を中心として、ウェルビーイングの哲学に関するオンライン研究会を実施した。②2020年12月にオンライン開催された第28回産業ストレス学会シンポジウム「これからの働き方を考える」で、鈴木貴之が「テクノロジー、幸福、朗働」という題目で発表を行った。③RISTEX松浦プロジェクトとの共催で、2021年1月に「思考力とウェルビーイング」と題したオンラインワークショップを開催し、批判的思考力や科学的思考力とウェルビーイングの関係、情報テクノロジーを用いてウェルビーイングを高める可能性などについて参加者で議論した。

### (3) 成果

今年度の到達点①：

（目標）人工知能研究の現状に関する理解を深める

実施項目1：人工知能研究の歴史と現状のレビュー（全体計画書の実施項目1）

成果：現在の人工知能研究が記号計算とニューラルネットのハイブリッド的性格を

強めていること、確率的なアプローチや機械学習が重要な要素となっていることなどが明らかとなった。

今年度の到達点②：

(目標) 現在の人工知能の可能性と限界の哲学的分析

実施項目2-1：人工知能の可能性と限界の解明（全体計画書の実施項目4）

成果：深層学習にもとづくニューラルネットは知能の本質に関する見方を大きく転換させる可能性がある反面、汎用知能を実現するためには記号計算的なアプローチとのハイブリッド化も必要となると考えられることや、人間の知能を再現することよりも、人間の知的可能性を拡張する道具としての可能性を追求する方が人工知能研究のプログラムとして興味深いことなどが明らかになった。

実施項目2-2：人工知能による徳の実現可能性（全体計画書の実施項目8）

成果：人工知能をはじめとする情報テクノロジーを倫理的徳や認識的徳を高めるための道具として利用する可能性があることや、その具体的事例とみなしうるものがすでに存在することが明らかになった。同時に、これらが悪徳の実現につながりうるという危険性も明らかになった。

実施項目2-3：情報テクノロジーによる人間知性の拡張可能性の検討（全体計画書の実施項目11）

成果：人工知能や機械を相互作用するアクターの一つとしてとらえ、人間と機械を相互作用するアクター同士のハイブリッドとしてとらえる視点が明確になり、人工知能の哲学が問題にすべきことがらとして、知性や認識をも含めた人間の諸技能・能力が人工知能との相互作用により変容すること（一概に、増強や拡張といったポジティブな側面だけに尽きない変化があること）が明らかになった。

実施項目2-4：人間知性と人工知能の協働可能性の検討（全体計画書の実施項目12）

成果：実施項目11の成果にもとづき、80年代の人工知能の哲学とは異なる問題系が明確化され、人間のよりよい生を実現するために人間と協働できるテクノロジーとはどのようなものかという問題設定が明らかになり、人工知能をどのようなツールとすべきか・するのが望ましいのかという技術哲学的観点からの検討の必要性が明らかになった。

実施項目2-5：ウェルビーイングに関する哲学的検討（全体計画書未記載項目）

成果：分析哲学におけるウェルビーイングの哲学の現状を確認するとともに、情報テクノロジーが発展した社会の望ましいあり方を考える上では、現在のウェルビーイングの哲学の枠組は不十分なものであり、われわれはどのような価値を重視すべきかを論じるための枠組が必要なことが明らかになった。

#### (4) 当該年度の成果の総括・次年度に向けた課題

新型コロナウイルス感染症の影響により、グラナダ大学の研究チームとの国際ワーク

ショップ開催など、いくつかの研究活動を実施することができなかった。その影響で、プロジェクト全体として進捗状況にやや遅れが生じている。今年度後半にはオンラインを前提として研究活動を再開することができ、研究会などではオンライン活動の利点も明らかになったため、次年度も引き続きオンラインを前提として活動を継続し、次年度後半までの活動によって当初の目標の達成を目指す。

### 2 - 3. 会議等の活動

| 年月日         | 名称   | 場所    | 概要  |
|-------------|--|-------|---|
| 2020年8月8日   | オンライン講演会   | オンライン | 一橋大学の久保明教氏が「ハイブリッドはいかに忘却されるか-現代将棋におけるソフトのツール化」という講演を行った。          |
| 2020年10月12日 | プロジェクト研究会  | オンライン | PHTR2020でのパネルセッションに向けて、提題者が発表を行った。                                |
| 2021年1月23日  | ワークショップ「思考力とウェルビーイング」                              | オンライン | 批判的思考力・科学的思考力とウェルビーイングの関係、情報テクノロジーを活用してウェルビーイングを高める可能性などについて議論した。 |
| 2021年3月20日  | Shiawase2021ワークショップ「哲学者とともに考えるAI時代のわたしたちのウェルビーイング」 | オンライン | 立花・植原・中澤の提題に基づいて、AI時代のウェルビーイングはどうあるべきかを議論した。                      |

### 3. 研究開発成果の活用・展開に向けた状況

2021年3月に、立花グループの立花・植原と染谷グループの中澤がテクノロジーとウェルビーイングをテーマとした一般向けのワークショップを開催した。

### 4. 研究開発実施体制

#### (1) 人工知能の哲学2.0の構築グループ（鈴木貴之）

東京大学大学院総合文化研究科

実施項目1：人工知能研究の歴史と現状のレビュー

グループの役割の説明：以下の実施項目を遂行するための準備作業として、文献調査や研究会の開催などを通じて人工知能研究の現状について引き続きレビューを行う。

実施項目2-1：人工知能の可能性と限界の解明

グループの役割の説明：汎用人工知能や自律的ロボットを実現するうえでの本質的な問題がどこにあるか、人間知性と人工知能にはどのような共通点と相違点があるかを明らかにする。具体的な活動としては、人工知能研究者・哲学者を講演者とした研究会の開催、人工知能研究者と人文科学研究者の対話を目的としたワークショップの開催などを行う。

実施項目3：成果物の作成準備

グループの役割の説明：本研究開発プロジェクトの主たる成果物となる出版物の構成や担当について、全体研究会で検討し、構想を確定する。

#### (2) 徳と人工知能グループ（立花幸司）

熊本大学大学院人文社会科学研究部

実施項目2-2：人工知能による徳の実現可能性

グループの役割の説明：徳を構成する各要素のどのような側面が人工知能によって実現可能であるのか、またどのような側面は実現が困難であるかを、道徳的徳と認知的徳の二つの観点から検討し、明らかにする。

#### (3) 拡張された心と人工知能グループ（染谷昌義）

高千穂大学人間科学部

実施項目2-3：人間知性と人工知能の協働可能性の検討

グループの役割の説明：人間の知的活動を人工知能がどのように補完あるいは補助し協働できるのか、望ましい協働はどのような姿を取るべきなのかを明らかにする。

## 5. 研究開発実施者

### 人工知能の哲学2.0の構築グループ（リーダー氏名：鈴木貴之）

| 氏名   | フリガナ    | 所属機関 | 所属部署       | 役職<br>(身分) |
|------|---------|------|------------|------------|
| 鈴木貴之 | スズキタカユキ | 東京大学 | 大学院総合文化研究科 | 准教授        |

### 徳と人工知能グループ（リーダー氏名：立花幸司）

| 氏名   | フリガナ    | 所属機関 | 所属部署       | 役職<br>(身分) |
|------|---------|------|------------|------------|
| 立花幸司 | タチバナコウジ | 熊本大学 | 大学院人文社会科学部 | 准教授        |
| 植原亮  | ウエハラリョウ | 関西大学 | 総合情報学部     | 教授         |

### 拡張された心と人工知能グループ（リーダー氏名：染谷昌義）

| 氏名   | フリガナ     | 所属機関   | 所属部署     | 役職<br>(身分) |
|------|----------|--------|----------|------------|
| 染谷昌義 | ソメヤマサヨシ  | 高千穂大学  | 人間科学部    | 教授         |
| 柴田崇  | シバタタカシ   | 北海学園大学 | 人文学部     | 教授         |
| 上杉繁  | ウエスギシゲル  | 早稲田大学  | 創造理工学部   | 教授         |
| 中澤栄輔 | ナカザワエイスケ | 東京大学   | 大学院医学研究科 | 講師         |

## 6. 研究開発成果の発表・発信状況、アウトリーチ活動など

### 6-1. シンポジウム等

| 年月日        | 名称  | 主催者  | 場所                     | 参加人数 | 概要   |
|------------|---|--|------------------------|------|--|
| 2020年11月6日 | Panel Session:<br>Artificial Intelligence as a Tool | Philosophy of Human - Technology Relations Conference 2020 | オンライン<br>(主催：トウウェンテ大学) | 約30名 | 人間の知性を拡張する道具としての人工知能の可能性について、鈴木・立花・柴田・上杉が発表を行い、その内容について参加者と議論した。 |

### 6-2. 社会に向けた情報発信状況、アウトリーチ活動など

#### (1) 書籍、フリーペーパー、DVD

- ・植原亮『思考力改善ドリル——批判的思考から科学的思考へ』、勁草書房、2020年

#### (2) ウェブメディアの開設・運営

- ・プロジェクトウェブサイト (URL : <http://updatingphilosophyofai.net/>、2019年2月開設)

#### (3) 学会 (6-4.参照) 以外のシンポジウム等への招聘講演実施等

- ・染谷昌義「身体性と運動性—キスの制御則に示される心のはたらき—」玉川大学応用脳科学研究センター「心の哲学研究研究部門」第14回研究会、2021年3月6日 (オンライン)
- ・柴田崇「サイボーグ思想における不便益の意義」第31回不便益システム研究会、2020年12月7日
- ・中澤栄輔「脳科学分野におけるELSI」応用脳科学アカデミーベーシックコース3「ELSI」第1回、NTTデータ経営研究所、2020年10月30日

### 6-3. 論文発表

#### (1) 査読付き (1件)

##### ●国内誌 (1件)

・柴田崇「メディア研究と心理学の接点：『探索モデル』」『新人文学部』北海学園大学大学院文学研究科、第17巻、50-67頁、2020年12月

##### ●国際誌 (0件)

#### (2) 査読なし (5件)

- ・鈴木貴之「深層学習の哲学的意義」『科学哲学』日本科学哲学会、53巻2号、pp.151-167、2021年3月（招待論文）
- ・植原亮「作り物の徳認識論の規範性」『情報研究』関西大学総合情報学部、第51号、pp. 1-20、2020年8月
- ・植原亮「人工知能は科学を人間から切り離してしまうのか？」『セミナー年報2020』、関西大学経済・政治研究所、pp.69-82、2021年
- ・染谷昌義「二元論の向こう側を探る自然学のプログラム」『現代思想』第48巻8号、pp.187-195頁、2020年6月
- ・柴田崇「AI vs. IA：論争に隠れた真の課題」『人文論集』北海学園大学人文学部、第70巻、115-126頁、2021年3月

### 6-4. 口頭発表（国際学会発表及び主要な国内学会発表）

#### (1) 招待講演（国内会議2件、国際会議0件）

- ・鈴木貴之「テクノロジー、幸福、朗働」第28回日本産業ストレス学会シンポジウム「これからの働き方を考える」、オンライン、2020年12月5日
- ・柴田崇「メディア研究と心理学の接点：『探索モデル』」公開シンポジウム「ネットメディアの生態心理学」日本心理学会第84回大会（録画）、公開期間：2020年9月8日～11月2日

#### (2) 口頭発表（国内会議0件、国際会議4件）

- ・Takayuki Suzuki. Two Concepts of Artificial Intelligence. Panel Session: Artificial Intelligence as a Tool. Philosophy of Human-Technology Relations Conference 2020. Online (University of Twente). 2020/11/6.
- ・Koji Tachibana. Artificial Intelligence as A Tool for Moral Education. Panel Session: Artificial Intelligence as a Tool. Philosophy of Human-Technology Relations Conference 2020. Online (University of Twente). 2020/11/6.
- ・Takashi Shibata. AI vs. AI: The Real issues hidden in the struggle. Panel Session:

Artificial Intelligence as a Tool. Philosophy of Human-Technology Relations  
Conference 2020. Online (University of Twente). 2020/11/6.

- Shigeru Wesughi. Considerations on Analysing Relations between Humans and AI Technologies Based on Archetypes of Instruments – Club-type and Pot-type. Panel Session: Artificial Intelligence as a Tool. Philosophy of Human-Technology Relations Conference 2020. Online (University of Twente). 2020/11/6.

(3) ポスター発表 (国内会議0件、国際会議0件)

#### 6-5. 新聞／TV報道・投稿、受賞等

(1) 新聞報道・投稿 (0件)

(2) 受賞 (0件)

(3) その他 (2件)

- 植原亮「人工知能(1) 不気味さの諸相と「汝自身を知れ」」『文部科学教育通信』、499号、pp.22-23、2021年(解説記事)
- 「人工知能(2) AIと人間にとっての創造性の意味」『文部科学教育通信』、500号、pp.22-23、2021年(解説記事)

#### 6-6. 知財出願

(1) 国内出願 (0件)

・

(2) 海外出願 (0件)

・