

Article

ア-ティ-クル

人工知能学会 倫理委員会の取組み

Introduction of JSAI Ethics Committee

松尾 豊
Yutaka Matsuo

東京大学大学院工学系研究科
Graduate School of Engineering, The University of Tokyo.
matsuo@weblab.t.u-tokyo.ac.jp

西田 豊明
Toyoaki Nishida

京都大学大学院情報学研究所
Graduate School of Informatics, Kyoto University.
nishida@i.kyoto-u.ac.jp

堀 浩一
Koichi Hori

東京大学大学院工学系研究科
Graduate School of Engineering, The University of Tokyo.
hori@computer.org

武田 英明
Hideaki Takeda

国立情報学研究所情報学プリンシプル研究系
Principles of Informatics Research Division, National Institute of Informatics.
takeda@nii.ac.jp

長谷 敏司
Satoshi Hase

SF 作家
Science-Fiction Writer.
haseo@white.plala.or.jp

塩野 誠
Makoto Shiono

株式会社経営共創基盤
Industrial Growth Platform, Inc.
m.shiono@igpi.co.jp

服部 宏充
Hiromitsu Hattori

立命館大学情報理工学部
College of Information Science and Engineering, Ritsumeikan University.
hatto@fc.ritsumeai.ac.jp

Keywords: ethics, ELSI, impact of AI, singularity, FLI, AI 100.

1. はじめに

本学会の倫理委員会を立ち上げる話が出たのは、2014年5月頃である。本稿の第一著者の松尾に、会長の松原仁氏から、倫理委員会を立ち上げることを検討してほしいという旨の話があった。名称やメンバ、活動内容などを検討し、理事会で認められたのが2014年9月^{*1}、第1回の倫理委員会が東京大学で開催されたのが、2014年12月であった。

倫理委員会のメンバには、本学会からだけでなく、学会外の有識者にも加わっていただくことが適切と考え、2

名の方に引き受けてもらった。倫理的な問題は社会全体に関わる問題であり、学会内の研究者だけで議論するのではあまりに視野が狭くなると考えたからである。一人は、人工知能に関する小説を多数書いており、また2015年の日本SF大賞を受賞されたSF作家の長谷敏司氏である。もう一人は、人工知能を含めたIT技術全般に詳しく、法律や経営にも造詣の深い(株)経営共創基盤取締役の塩野誠氏に引き受けていただくことができた。

学会側のメンバとしては、倫理的な問題について以前から意見を発信している元会長の東京大学堀浩一氏と京都大学西田豊明氏、またLODを中心に人工知能技術の普及に務めている国立情報学研究所武田英明氏、AIR (Acceptable Intelligence with Responsibility) 研究

会を主催している立命館大学服部宏充氏をお迎えすることができた。松尾を加えた合計7名で現在のところ委員会は構成されている(今後も必要に応じて委員の追加を行っていく予定である。なお、本稿は松尾が文責であるため、他の共著者には敬称を付している)。

倫理委員会という名称についてであるが、実はこの名称はかなり迷った末に付けられたものである。当初から、倫理委員会という名称にするか、あるいは例えば「人工知能と未来社会検討委員会」などの名称にするかを最後まで決めかねた(現在でも名称に関する議論は続いている)。この委員会の本来の趣旨は、人工知能研究あるいは人工知能技術と社会との関わりを広く捉え、それを議論し考察し、社会に適切に発信していくことである。社会との関わり

*1 実際には、本学会では2007年から倫理委員会という名称の委員会が設置されていたが、継続的な活動は行われておらず休眠状態であった。

りの中で、倫理というのは一つの側面に過ぎないから、「倫理委員会」と称するのは想定する役割よりもあまりに狭い^{*2}。

一方で、「倫理委員会」という名称はさまざまな組織ですでにつくられており、何らかの影響をもつ「怖い」委員会であるというイメージもある（実際、「イブの時間」というロボットと人間が入り交じる未来社会を描いたアニメーション作品[吉浦 10]では、倫理委員会なる組織が登場し、ロボットが人間のような行動をすることを取り締まっている）。倫理委員会という名称の組織をつくるということは、人工知能に関する意識を高めなければならないことの学会としての表明でもあり、その結果としてこうした委員会が、社会に対して何らかの発信力をもつことになる。このことは、倫理委員会が社会と対話しながら活動していくうえで重要である。したがって、その意図する本来の役割と名称に齟齬があることは理解しつつも、社会から見たときにその役割が第一近似としてでもわかりやすい、「倫理委員会」という名称とすることにした。

こうした名称を選んだせいもあるが、倫理委員会は、設立当初から反響が大きかった。まだ第1回の会合しか開催していない時点で、新聞社から複数の取材の依頼があり、日本経済新聞や読売新聞などで取り上げられた[日経 14, 読売 15]。また、2015年2月に開催された第2回の倫理委員会には、NHKのカメラが入り、2015年2月19日「おはよう日本」の一つのコーナーである、ビジネス最前線「進化する人工知能」において、委員会の活動が発足したことについても放送された。

第一著者である松尾自身の思いとしては、今回の倫理委員会のような活動の必要性を感じたのは、昨年の「表紙

問題」であった。2014年1月に本学会の学会誌の名称および表紙を変更したときに、社会的に大きな反響があった。その経緯や議論については、2014年3月号の小特集[栗原 14]、ならびに2015年1月号の坊農氏のコラム[坊農 15]と新しい表紙についての記事[大澤 15]に詳しい。当時は、人工知能が社会に対してこれほどまでにインパクトをもっていることを十分に想像していなかった。そして、人工知能のプレゼンスが高まる中で、社会との接点を真剣に考えないといけない段階に来ていることを痛感した。今回、倫理委員会という活動の場を与えていただき、光栄に思っている。

本稿では、まだ始まったばかりではあるが、倫理委員会がどのような活動を行っているかを紹介するために、現在までの議論について、関連の取組みについて、また今後の取組みについて述べたい。

2. 第1回倫理委員会での議論について

第1回の倫理委員会は、2014年12月15日、東京大学にて開催された。倫理委員会の委員に加え、本誌編集委員会からは、編集委員長の栗原 聡氏、および副編集委員長の山川 宏氏が参加した。さまざまな話題を移り変わりながら議論が進展したが、その内容をまとめたものが以下である。

2.1 正しい現状理解の必要性

まず、議論の土台として現状理解を正しくする必要がある。シンギュラリティをめぐる議論では、人間の知能を大きく上回る人工知能が出現する、それが人間のコントロールを離れて暴走するかもしれないなどといった極端な議論が見られるが、研究者から見ると信ぴょう性は疑問が残る。シンギュラリティで議論されているような「真に自己を設計できる人工知能」の実現は、技術的にははるかかなたにあり、現在のところその糸口すらつかめていないことは正しく理解する必要がある。

一方で、科学技術が進むことで我々自身の知能が以前よりも大幅に「増幅」されていることは事実である。そして、「人間の知能を超える知能」は、人間と

機械が協調することで実質的にすでに生み出されている。PCや検索を始め、さまざまな技術により、知的活動を効率的に行うことができる。また、電力や銀行などのインフラを含め、社会全体のなかに機械が組み込まれ、機械と人間が入り混じる社会になってきている。この傾向はますます進んでおり、こうした変化に対して、人間は柔軟に適応している。

すなわち、現状で、すでに人間社会に機械は入り込んでおり、人間がこれまで行ってきたものを大きく超える知的活動が行われるようになる変化はすでに起こっている。一方で、人工知能が人工知能をつくるというような技術ははるかかなた先であることも理解しなければならない。これがひとまずの議論の土台である。

2.2 我々の役割の明確化

では、我々、倫理委員会の役割は何だろうか。人工知能が人工知能をつくるというような未来は当然来ないので、特に何もしなくてよいのだろうか。いや、そうではないだろう。往々にして、専門家は自らの技術を過小評価しがちである。人工知能研究者にとってみると当たり前の技術が社会に大きなインパクトを与えてしまうこともあり得る。今後、人工知能の技術が、さまざまな形で社会のインフラに組み込まれていくことは確実であろうから、そうしたときに、思いがけず大きなインパクトを与えてしまうことがあるかもしれない。

専門家として、予見できるものを予見しておくことは、社会に対する誠実な態度であろう。さまざまな諸問題が起こらないうちに、どういう問題が発生する可能性があるのか、それに対して我々はどうのような解決策をもつことができるのかなどを議論しておく必要があるだろう。例えば、遺伝子工学における例として、クローン技術が生まれたあと、人間のクローンをつくってはいけないと禁止された（2000年に交付された「ヒトに関するクローン技術などの規制に関する法律」によって禁止されている）。人工知能技術に関しても同じような禁止すべき行為があり得

*2 倫理よりも、ELSI (Ethical, Legal and Social Issues (あるいは Implications)) のほうが近いであろう。しかし、ここでは、技術の未来予測やどのような方向に研究が進んでいくべきかも含めた議論を扱いたいと思っており、ELSIよりもさらに広い概念かもしれない。

るのかもしれない。例えば、後述するAAAIのパネルでは、「犯罪的なAI、軍事AI、中毒や依存をもたらすようなAIに基づくシステム」に関しては、社会的な注意を払うべきであると議論されている。もしこのようなAIを抑制すべきことが適切なのであれば、社会を巻き込んで（ときには国際社会を巻き込んで）いくことも必要であろう。あるいは、禁止すべき行為を特定するのが難しい場合には、倫理規範、行動規範といった形で研究者や開発者の行動を良い方向に向かうように企図する必要があるのかもしれない。

また、東日本大震災における原子力発電所の事故を鑑みるに、考え得る最悪なシナリオとその対応を列挙することも専門家の重要な役割であろう。さらには、現在の技術だけでなく、今後の技術の方向性を指し示すことも社会と対話するうえで重要な役割であろう。

2・3 考え方の指針

では、こうした役割を果たしていくために、どのような考え方が活動の指針となるだろうか？

まず、我々は社会と対話しなければならない。倫理観は研究者自身ではなく、社会全体でつくっていくものである。そして、人々の役に立つ人工知能、人々の生活を幸せにし、人生をより豊かにするような人工知能、つまり「万人のための人工知能」を目指すべきであろう。これは科学技術一般にいえることであろうが、ともすれば「一部の人が活用できる人工知能」となりがちであるのを、「万人のための人工知能」であると宣言するのは重要なことである。

では、人々を幸せにし、人生をより豊かにするという事はどういうことであろうか。人間にはさまざまな価値基準があるので、一般的には、ある特定の尺度でもって良い悪いを判断することは難しいだろう。人工知能が使われるうえで、さまざまな良い点や悪い点が出てくるであろうが、おそらく多くの人にとって社会における人工知能システムの良し悪しを判断する価値基準の一つが、「人間としての尊厳が守られるか」ではないだろうか。人工知能

が発展することで、人間としての尊厳が侵される（例えばコンピュータに指示されるとおりに動く生産性が上がるので、それを強要される）ようなことは不幸なことである。人工知能の技術は、「人間の尊厳」を守るように使われていくべきではないだろうか。

また、人工知能が社会の期待に沿って倫理的に使われるためにはどのような手段を講じれば良いだろうか。そのためには人工知能システムがオープンであること（透明性をもつこと）が必要であるし、また中で用いられているアルゴリズムがブラックボックス化するのではなく、説明可能であることも必要であろう。もし人工知能が「暴走する」という危惧を多くの人がもつのであれば、その制御権を複数の人間（市民）に分散することなども必要かもしれない。

こうした議論を通じて明らかになってくるのは、人工知能の倫理的問題というときに主に議論すべきは、ロボット三原則のような「人工知能がもつべき倫理」ではなく、人間の側の倫理、すなわち「人工知能を使う人間の倫理」、「人工知能を開発する人間の倫理」であるということである。

2・4 職業の問題

人工知能が職業を奪うという話がよく聞かれる。オックスフォード大学のOsborne氏らの論文によると、702の職種のうち47%が今後10年以内になくなる確率が高いとされている[Frey 13]。この真偽については賛否両論あるだろうが、こうした議論に社会の関心が高いことも事実である。

職業の問題は、より抽象化しているという点、人工知能の普及が短期にもたらす社会的あるいは個人への顕著な影響である。科学技術による社会的変化が起こること自体は、飢餓が減り、病気が減り、より生産性が上がり、人類全体の効用を上げるものであろうから、マクロには良いことであるとしても、それが急激に起こることで、短期的には社会や個人に何らかの悪影響を及ぼすこともあり得る。産業革命以後の1811年頃から、イギリスにおいて、ラッドイト運動と呼ばれる運動が起こり、労

働者らが機械普及による雇用の喪失を懸念して機械を破壊する運動を行ったのは有名であるが、人工知能の技術の普及においても、こうした急激な変化がもたらすかもしれない悪影響については十分な配慮をすべきであろう。

こうした問題は、より突き詰めていけば、がんばって勉強し、夢の職業に就いたときにその職業が大きく変化しているというようなことがあってよいのか、あるいは、自分の子供に大きな教育投資をした人が、人工知能の普及の影響により報われないということがあってよいのかといった問題を内包する。何を善とし社会として奨励するのかという問題とも密接に関わる問題である。

また、人工知能による生産性の向上は、富の再分配の議論とも関わる[井上13]。「機械との競争」[ブリニョルフソン13]の中でも書かれているように、資本家と労働者のもち得る富に偏りが大きくなってくると、これをどう再分配するかは社会全体での問題である。技術革新が早過ぎることへの短期的な影響についての社会全体での議論は必要であろう。

2・5 人工知能の「心」の問題

ここまでの議論は、科学技術全般に当てはまるものが多かった。特定の科学技術が急速に進展するときに、どのような社会的影響を及ぼすのか、専門家集団はどのように社会と対話すべきかという一般論として、研究領域を問わず、上記のような議論を行っていく必要があるだろう。

一方で、人工知能に特有の問題はあるだろうか。例えば、心の問題というのは、人工知能特有の大きな問題の一つかもしれない。心は生命と同等あるいはそれ以上に人間の本質を占めている。我々の心が相互依存であることを鑑みれば、心をもつ人工知能システムをつくってもよいかどうかというのは大きな論点である。ここで「心をもつ」という場合の「心」が何を意味するかは難しい問題であるが、ここでは単純に、振舞いとして「心をもっているように見える」人工知能システムを、心をもつ人工知能システムと呼ぶことに

しよう。

心をもつ人工知能システムは、例えば、人間が恋愛感情を抱いてしまう恐れがある(映画「her / 世界で一つの彼女」ではこの世界が描かれている)。最近のゲーム(例えば、ラブプラスなど)の進展やソーシャルゲームにおける課金の問題などを合わせて考えると、心をもつ人工知能システムが悪用されることも十分に想像できる。さらに、人間が感情移入することで、例えば、そうした人工知能プログラムを停止させてよいか、プログラムを壊してよいかなど、人間の心にもつわるさまざまな問題を誘発する可能性がある。

ここでの「心の問題」は一例であるが、こうした人工知能に特有の問題について、どのような危険性があるのか、それに対してどのように対応していくことができるのかを考えていかなければならない。

3. 第2回倫理委員会での議論について

第2回の倫理委員会は、2015年2月9日、東京大学にて開催された。編集委員会から、編集委員長の栗原 聡氏が参加した。

第2回の委員会では、まず、諸外国における関連の取組みについて紹介された。特に、その中でも主要な動きである、FLI (Future of Life Institute) や、StanfordのAI 100 (One Hundred Year Study of Artificial Intelligence) などの活動について共有された。いずれの動きも、社会に対して、レターやホワイトペーパーなどの形で情報を発信している。その内容については、次章に関連動向としてまとめて述べる。

次に、倫理委員会がこのような対外的な声明を発信していくべきかどうか、発信する場合にはどのようにすべきかに関しての議論が行われた。その論点について以下にまとめる。

3.1 誰に向けての声明か

倫理委員会が、将来的に対外的な声明を発信することを考えたとき、その声明が誰に対してのものかを考えることは重要である。倫理委員会から出す声明は倫理の面にとどまらず、広く社会への影響を考慮すべきものである。

本来、倫理委員会の活動は、倫理に限らず広い範囲を対象としていることを考えると、声明もより広い範囲であるべきである(例えば、AI 100のホワイトペーパーには、倫理や法律など18項目に関しての議論が掲載されている)。

そうした際に、まず問題を提起するために研究者に向けてメッセージを発信するのか、あるいは、最初から(人工知能研究者を含む)市民に向けて出すという考え方がある。研究者内部での意見を統一してからという考えもあるが、こうした問題は、社会と対話しながらその方向性をくり上げていくものであろうから、最初から市民に向けた声明を発信していくのが適切ではないだろうか。

3.2 声明を出す目的

次に、声明を出す際には、その目的に関して明確に認識しておかねばならない。こうした声明を出す目的は、人工知能が社会に対してもつ見えない影響に対して、研究者自らが、あるいは市民が意識的になることである。人工知能は見えないものであり、それに気付かせる、あるいは気付いてもらう必要性は大きい。

研究者には、自分達が研究している内容が、かなり近いところで社会に影響を及ぼす可能性があるのだということに自覚的になってもらい、こうした問題を積極的に考えてもらいたい。そしてこの影響を意識して研究活動を行ってもらいたい。さらにいえば、社会に良い影響を与える研究活動を積極的に推進してもらいたい。また、市民にとっては、今後、人工知能の発展に伴って、どういう可能性やリスクがあるのかを正しく理解し、それに対してどのような選択肢があり得るのかを考えてもらいたい。

実は、こういった技術の社会的影響に関する懸念は、人工知能に限らず、情報技術全般(例えば、ソーシャルメディアやスマートフォンの利用)に当てはまるものであろう。ソーシャルメディアやスマートフォンの利用の問題についてもマスコミで取り上げられていることを目にすることも多い。我々は、特に、人工知能技術の影響に対して、

注意深く情報を発信していくべきである。

3.3 社会への影響について

では、人工知能が広まることで起こり得る社会への影響とは例えばどんなものであろうか。昨今、社会で必要とされる知性の質が変わっている。情報に接する機会が増え、日々、大量の情報にさらされる中で、我々自身の能力としては、例えば、ゆっくり考える能力よりも、大局的な情報に基づいて瞬時に判断する能力などが必要とされるようになってきているのかもしれない。知性の質が変わることは人間の尊厳に直接関わる。今まで高く評価されてきた知性を必要とする仕事が、そうでなくなることもあり得るからである。

今後、例えば、人間の共感性や心が必要とするような仕事に重点が置かれるようになるのかもしれない。人間が「機械的な」仕事から解放され、より人間的な共感性や心が必要とする仕事に、自ら主体性をもって取り組むことができるようになるかもしれない。こうした変化は、長期的には「やりがいのある仕事」を提供し、人間の「生きがい」につながり、社会をより豊かにするかもしれない。そのように技術と社会が進んでいくことを志向することも、一つの方向性であろう。

グローバルに人工知能技術が発展していくなかで、日本における人工知能という視点も重要である。日本は古来から自然と調和してきた国であり、さまざまなものが共生する文化をつくり上げている。「ドラえもん」や「鉄腕アトム」のように、人間と機械が共生し、人間に機械が寄り添ってくれる良いイメージがすでにつくり上げられている。人間と機械が共生しながら、人間がより人間らしく生きられる社会をつくり上げていくような世界観を日本から発信していくこともまた重要な方向性であろう。

3.4 シンギュラリティについて

倫理委員会の中でも、技術に対する見方はさまざまである。例えば、「シンギュラリティ」が起きるかどうかについてすら、議論が分かれた。真に自己設計できる機械を(例えば今後10年で)

つくれる見込みはないとする立場と、真に自己設計できるかどうかは関係なく、誰も理解できない、誰も制御できないスーパーインテリジェンスの出現こそが脅威であるという立場などがある。

一方で、時間の制約を外し、(今後10年ではなく)100年という期間で見れば、シンギュラリティが起きる可能性があることは、共通の合意であった(シンギュラリティについては、情報処理学会の2015年1月号での特集[塚本15]も参考にされたい。例えば堀氏は、社会から見た人工知能の不安について議論している[堀15])。

このように、研究者の統一的な見解を出すことは容易なことではないが、今後、声明を出していくうえでは、何らかの見解をベースに、社会的な影響を議論していく必要がある。その際にも、短期の議論と長期の議論、あるいは期待値的な見通しと悲観的な見通しは分けるべきとする意見があった。

4. 関連の取組み

4.1 海外の取組み

本学会の倫理委員会は、2014年の春頃から検討を始め、9月に立ち上がったものであるが、時期を同じくして、海外でもいくつかの活動が立ち上がっている。そのうちの有名な二つを紹介しよう。

一つは、**The Future of Life Institute (FLI)** というボランティアベースの団体である。ボストンを活動の中心とし、2014年3月から活動が開始された。MIT教授の**Max Tegmark**氏、スカイプの創業エンジニアである**Jaar Tallinn**氏、UCサンタ・クルーズの**Anthony Aguirre**氏ら5人が中心となつてつくられたものである。人類の直面するリスク、特に人工知能のもたらす影響についての研究とアウトリーチ活動を行っている。

アドバイザーボードには、著名人が並び、**Superintelligence** [Bostrom 14]の著者であるオックスフォード大学の**Nick Bostrom**氏や、映画俳優の**Morgan Freeman**氏、宇宙物理学者の**Stephen Hawking**氏、**SpaceX**

やテスラモーターズの創業者である**Elon Musk**氏、UCバークレーの教授で人工知能の教科書で有名な**Stuart Russell**氏らが含まれる。なかでも、**Elon Musk**氏は2015年初頭に同団体に1,000万ドル(約12億円)を寄付したことが話題になった。

FLIのサイト^{*3}において、“**Research Priorities for Robust and Beneficial Artificial Intelligence: an Open Letter**”(頑健で有益な人工知能への研究の優先順位: オープンレター) というものが掲載されており [Russell 14]、これに同意する6,000名もの人の中には、著名研究者も散見される(日本からも何人かサインしている)。この中で、短期的な研究の優先度について、および長期的な研究の優先度について述べられており、倫理委員会とも関連する議論が繰り返されている。

一方、このFLIと並んで有名なものが、Stanfordの**One Hundred Year Study on Artificial Intelligence (AI 100)**という活動である。2007～09年までAAAIの会長を務めた**Eric Horvitz**氏(現マイクロソフト)らを中心とした活動である。その在任期間中に、**Long-term AI Futures**というパネルを企画し [Horvitz 09]、**Tom Mitchell**氏や**David Waltz**氏、**Peter Stone**氏、**Andrew Ng**氏や**Sebastian Thrun**氏、日本からは、石田亨氏など24人をメンバとして、人工知能技術の将来的な社会的インパクトに関する議論が行われた。

この活動がベースとなり、スタンフォード大学がホストする形で、2014年12月に、**Eric Horvitz**氏とバイオ工学と計算機科学の教授である**Russ Altman**氏が、この活動を立ち上げた。初期メンバには、**Barbara Grosz**氏、**Deirdre Mulligan**氏、**Yoav Shoham**氏、**Tom Mitchell**氏、**Alan Mackworth**氏らが加わっている。そのサイト^{*4}には、次のように記されている。

「人工知能は、科学における最も深

遠な企図の一つであり、人間生活のすべての側面に影響を与えるものである。」こう述べたのは、本活動を支援していただいたスタンフォード大学の学長**John Hennessy**氏である。「スタンフォードのAIにおける開拓者としての役割と領域にこだわらない思考方法を踏まえると、人工知能がどのように我々の子供や子供の子供に影響をあたえるかという議論をホストすることは、義務であると同時に、我々にはその資格があるとも感じている。」

AI 100からはホワイトペーパーが出され、倫理面だけでなく、法律、経済、幸福や自由、犯罪、機械との協調など18項目にわたって議論されている [Horvitz 14]。

そのほか、ケンブリッジ大学の**Center for the Study of Existential Risk (CSER)**、オックスフォード大学の哲学科におかれた**Future of Humanity Institute (FHI)**、また、シリコンバレーにあるNPO機関である**The Machine Intelligence Research Institute (MIRI)**などでも関連した活動を行っている。こうした取組みについては、江間氏の文献にまとめられている [江間 15]。

4.2 国内の取組み

一方、国内でもさまざまな取組みが始まっている。

本学会では、2014年度の全国大会において、ダウンゴ人工知能研究所の山川宏氏(本誌副編集委員長)が中心となり、「映画「トランセンデンス」共同特別企画～エヴリンの夢～」というセッションが開かれ、映画「トランセンデンス」で描かれているようなAIの未来像について議論が行われた。2045年問題に関する著書 [松田 12] を執筆している神戸大学の松田卓也氏も登壇し、また本委員会の堀氏も議論に加わって、大変盛り上がったと聞いている。このセッションで、初めて人工知能に関する未来社会の話聞いた方も多いのではないだろうか。

このセッションの企画の母体となったのが、山川氏や国立情報学研究所の

*3 <http://futureoflife.org/>

*4 <https://ai100.stanford.edu/>

市瀬龍太郎氏（本誌シニア編集委員）を中心とする、AGI（Artificial General Intelligence：人工汎用知能）研究会である。AGIやシンギュラリティに興味のある研究者で構成され、それに関する書籍や論文の調査や議論を2013年から継続的に行っている（AGIやシンギュラリティの解説については[山川 15]を参照されたい）。また、2015年の3月からは、早稲田大学の井上智洋氏らを中心に、AI社会論研究会が3月から活動を開始している。日本シンギュラリティ協会理事の佐々木健美氏も参加していると聞いている。

一方、関西を主な活動の中心として、京都大学（本年4月から東京大学）の江間有沙氏、立命館大学で本委員会委員の服部氏らが2014年に立ち上げたAIR（Acceptable Intelligence with Responsibility）研究会では、すでに3回ほどワークショップを開催しており、その報告がWebに公開されている*5。「人工知能が浸透する社会を考える」ことをテーマとしながら内容の濃い議論を行っており、本誌の2014年9月号には特集「人工知能が浸透する社会を考える」も掲載された[服部 14]。

こうした最近の動きとは別に、古くから本学会では社会との役割について考えられてきた。本学会の第二種研究会である「社会におけるAI研究会」（SIG-SAI）は、2006年から社会と人工知能の関わりについて議論を行っている。慶應義塾大学の神成淳司氏、現本学会会長の松原 仁氏、産業技術総合研究所の野田五十樹氏らが中心に活動を立ち上げた。当時は現在ほど人工知能がブームではなかったため、人工知能をいかに社会で活用していくかが中心のテーマであった。

省庁や国の機関でもさまざまな取組みが行われている。JSTにおいて2年ほど前から、人工知能を主要なテーマとする「知のコンピューティング」という活動が立ち上がり、CREST事業にもつながっている。これに関連し、倫

理や社会的影響について議論する「知のコンピューティングとELSI/SSH*6」ワークショップが2014年9月に開かれた。また、同じくJSTが主催し、2014年11月には、「サイエンスアゴラ2014 激論！ 先端ICTの光と影」というワークショップも開かれている。

総務省では、2015年2月から「インテリジェント化が加速するICTの未来像に関する研究会」を開催している。シンギュラリティをテーマにICTの未来を考えるものであり、省庁が初めて「シンギュラリティ」をテーマにした委員会を開催したと話題になった。また、経済産業省においても、2014年から行われている「情報経済小委員会」などで人工知能に関する議論が行われており、今後、経済産業省においても、人工知能の技術的な発展のみならず、産業構造全体への影響、雇用や生産性に関する影響、社会システム全体に関する影響についても議論が進んでいくものと思われる。

5. 今後の展望

本稿では、倫理委員会の立上げの経緯と第1回、第2回の委員会での議論について、および、国内外での関連の取組みについて述べた。

次回となる第3回の倫理委員会は、2015年6月1日の午後、はこだて未来大学において開かれる本学会全国大会の中で、公開セッション（一般の方にも公開される）として開催される予定である。興味のある方は、ぜひ参加いただき、議論に加わっていただきたい。

今後の活動としては、こうした委員会を継続していくことで、議論を深めると同時に、適切なタイミングで対外的な声明などの形で情報発信を行っていきたいと考えている。さらに、国内での関連の動きと連携していくと同時に、FLIやAI100などの国際的な組織とも連携し、日本からも積極的に議論に加わっていきけるような体制を模索したいと考えている。

今後、人工知能技術が進展していくなかで、研究や技術のレベル自体をますます向上させ、国際的な学術コミュニティでの日本の研究のプレゼンスを

向上させることは重要である。また、それを産業と連携させて、産業競争力につなげていくような取組みも不可欠であろう。同時に、本委員会でも意図するような、社会における影響を考え、人工知能技術が人々の幸せや夢の実現に寄与し、人工知能の発展が社会全体から歓迎されるように、研究コミュニティと社会全体の正しい関係をつくっていくことも必要な活動ではないかと考えている。本会の会員の皆様には、ぜひこうした趣旨に賛同いただき、ご支援、ご指導を賜れば幸いである。

謝 辞

倫理委員会の活動を支援していただいている、本学会会長のはこだて未来大学松原 仁氏、前会長の慶應義塾大学山口高平氏、編集委員長栗原 聡氏、副編集委員長の山川 宏氏、広報担当理事の津本周作氏、海外活動の調査でご協力いただいた京都大学（本年4月から東京大学）の江間有沙氏、議事運営に協力いただいている東京大学小川奈美氏の諸氏に感謝いたします。

◇ 参 考 文 献 ◇

- [坊農 15] 坊農真弓：表紙に込めたメッセージ：女性とAI、人工知能、Vol. 30, No. 1, p. 34 (2015)
- [Bostrom 14] Bostrom, N.: *Superintelligence: Paths, Dangers, Strategies*, Oxford University Press (2014)
- [ブリニョルフソン 13] エリク・ブリニョルフソン、アンドリュー・マカフィー：機械との競争、日経BP社 (2013)
- [江間 15] 江間有沙：「人工知能と未来」プロジェクトから見る現在の課題、第29回人工知能学会全国大会 (2015)
- [Frey 13] Frey, C. B. and Osborne, M. A.: The future of employment: how susceptible are jobs to computerisation?, <https://ai100.stanford.edu/reflections-and-framing> (2013)
- [服部 14] 服部宏充、江間有沙：特集「人工知能技術が浸透する社会を考える」にあたって、人工知能、Vol. 29, No.5, pp. 480-481 (2014)
- [堀 15] 堀 浩一：シンギュラリティへ向けてあなたと私はどうしたいか？、情報処理、Vol. 6, No. 1, pp. 41-43 (2015)
- [Horvitz 09] Horvitz, E. and Selman, B.: Interim report from the panel chairs, *AAAI Presidential Panel on Long Term AI Futures* (2009)
- [Horvitz 14] Horvitz, E.: One Hundred Year Study on Artificial Intelligence: Reflections and Framing, Oxford

*5 <http://web4ais.wpblog.jp/>

*6 SSHは人文社会科学（Social Sciences and Humanities）の略。

University Research Paper (2014)

- [井上 13] 井上智洋：技術的失業 私たちは機械に仕事を奪われていくのだろうか，エコノミスト別冊，Vol. 91, No. 56 (2013)
- [栗原 14] 栗原 聡，松尾 豊：「人工知能」表紙問題における議論と論点の整理」にあたって，人工知能，Vol. 29, No. 2, p. 166 (2014)
- [松田 12] 松田卓也：2045年問題 コンピュータが人類を超える日，廣済堂新書 (2012)
- [日経 14] 人工知能の未来を注視し研究を怠るな，日本経済新聞 (2014年12月21日)
- [大澤 15] 大澤博隆：2015年表紙更新にあたって—前年の「表紙問題」のまとめとこれから—，人工知能，Vol. 30, No. 1, pp. 2-6 (2015)
- [Russell 14] Russell, S., et al.: Research priorities for robust and beneficial artificial intelligence: An open letter, *Future of Life Institute* (2014)
- [塚本 15] 塚本昌彦 ほか：人類と ICT の未来：シンギュラリティまで30年？，特集，情報処理，Vol. 56, No. 1 (2015)
- [山川 15] 山川 宏，市瀬龍太郎，井上智洋：汎用人工知能が技術的特異点を巻き起こす，信学誌，Vol. 98, No. 3, pp. 238-243 (2014)
- [読売 15] 人工知能 急速に普及—一人の尊厳社会で議論—，読売新聞，2015年1月11日
- [吉浦 10] 吉浦康裕，太田優姫：イヴの時間，スクウェア・エニックス (2010)

2015年4月13日 受理

— 著者紹介 —



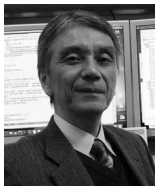
松尾 豊 (正会員)

1997年東京大学工学部電子情報工学科卒業。2002年同大学院博士課程修了。博士(工学)。産業技術総合研究所，スタンフォード大学を経て，2007年10月より，東京大学大学院工学系研究科准教授。2012～14年人工知能学会編集委員長，2014年より本学会倫理委員長。専門は，Webマイニング，Deep Learning，人工知能。



西田 豊明 (正会員)

1977年京都大学工学部卒業。1979同大学院修士課程修了。1993年奈良先端科学技術大学院大学教授，1999年東京大学大学院工学系研究科教授，2001年東京大学大学院情報理工学系研究科教授を経て，2004年4月京都大学大学院情報学研究所教授，現在に至る。会話情報学，原初知識モデル，社会知のデザインの研究に従事，本学会会長 (2010～2011年度)，日本学術会議連携会員 (2006年～)，情報処理学会フェロー，電子情報通信学会フェロー，JST CREST「共生社会に向けた人間調和型情報技術の構築」研究領域総括 (2014年～)，AI & Society 誌 Associate Editor。



堀 浩一 (正会員)

東京大学大学院工学系研究科教授。2008～10年本学会会長，2010年～本学会顧問。知識処理システム構築のための知識獲得の困難に直面したのち，研究すべきは，知識の獲得ではなく知識をつくり出すことの支援であると考えようになり，創造活動支援システムの研究を行ってきた。人間がもっと創造的に仕事するための道具としての人工知能の研究を行ってきたが，単なる道具では済みそうになってきたので，最近では人工知能倫理の問題にも興味をもっている。



武田 英明 (正会員)

1986年3月東京大学工学部卒業。1988年3月同大学院工学系研究科修士課程修了。1991年3月同博士課程修了。工学博士。ノルウェー工科大学，奈良先端科学技術大学院大学を経て，2003年5月より国立情報学研究所教授。2006年4月～2010年3月同学術コンテンツサービス研究開発センター長，2005年12月～2010年3月東京大学客員・特任教授。知識共有，セマンティック Web などの研究に従事，AAAI, Design Society, 情報処理学会各会員。



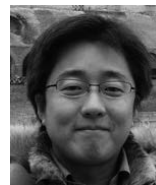
長谷 敏司

1996年3月関西大学文学部卒業。2001年角川スニーカー大賞にて作家デビュー。2009年よりSFを本格的に書き始め，2009年『あなたのための物語』，2012年『BEATLESS』で日本SF大賞ノミネート，2014年『My Hmanity』で日本SF大賞受賞。作品は近未来を舞台にAIが登場するものが多く，2014年ニコニコ学会参加をきっかけに，AIが現実社会に実装されるとき，社会の変化を望まない人々にまで触れてしまうインパクトについて考えるようになる。日本SF作家クラブ会員。



塩野 誠

1998年慶應義塾大学法学部政治学科卒業。2008年ワシントン大学ロースクール法学修士。株式会社経営共創基盤取締役。国内・海外にて企業，政府機関に対して戦略立案・実行のアドバイスをを行っている。テクノロジー関連の事業開発，企業提携やM&A，企業危機管理の実績を多く有する。



服部 宏充 (正会員)

2004年3月名古屋工業大学大学院博士課程修了。博士(工学)。2004年4月～2007年3月日本学術振興会特別研究員(PD)，リバプール大学，およびマサチューセッツ工科大学客員研究員を経て，2007年4月～2014年3月京都大学大学院情報学研究所助教。現在，立命館大学情報理工学部 准教授。マルチエージェントシステム，社会シミュレーション，デザインに関する研究に従事。本学会の「表紙問題」における社会の反応から，人工知能と社会の関係に関心をもち，有志と共にAIR (Acceptable Intelligence with Responsibility) の活動を開始。