

本ニュースレターでは2015年9月から11月までのAcceptable Intelligence with Responsibility (AIR) 研究グループの活動を報告する。

## 第5回ワークショップ報告

9月3日(木)に京都の国際高等研究所で第5回ワークショップを行った。7月に開催した第4回ワークショップを踏まえ、インタビュー調査とアンケート調査の具体的な内容を議論した。インタビュー調査はオーラルヒストリーの一次資料となることを目指すことで合意した。アンケート調査は人工知能学会「レクチャーシリーズ」を参照しながら、人工知能研究の多様性を捉える尺度を作成するための議論を行った。

本ワークショップは公益財団法人国際高等研究所の平成27年研究プロジェクトの助成を受け、秋谷直矩氏(エスノメソドロジー)、市瀬龍太郎氏(機械学習)、大家慎也氏(テクノロジーの政治・倫理)、大澤博隆氏(ヒューマン・エージェント・インタラクション)、大谷卓史氏(情報倫理学)、神崎宣次氏(倫理学)、久木田水生氏(哲学・倫理学)、久保明教氏(文化人類学)、服部宏充氏(マルチエージェントシステム)と江間有沙(科学技術社会論)の10名が参加した。



### 土屋俊氏インタビュー

9月18日(金)に筑波大学東京キャンパス文教科舎で土屋俊氏(大学評価・学位授与機構教授)にインタビューを行った。本インタビューデータはHPなどで公開する予定である。

本インタビューは国際高等研究所の助成を受け、秋谷氏、大澤氏、神崎氏、久木田氏、服部氏と江間の6名が参加した。



## 第6回ワークショップ報告

2014年から活動を継続する中で、人工知能と社会の関係について異分野間対話を試みている多くの人たちとの出会いがあった。そこで去年に引き続き科学技術社会論学会にて公開ワークショップを開催することで、彼らに話題提供いただき交流する場を企画した。

### ワークショップ概要

11月21日(土)の13時20分より東北大学川内南中キャンパスにて公開ワークショップを開催した。ワークショップには、次ページに紹介する企画者4名(AIR)と外部の方5名に話題提供いただいた。AIRメンバーからは神崎宣次氏(倫理学)、久木田水生氏(哲学・倫理学)、西條玲奈氏(分析哲学)、田中幹人氏(科学技術ジャーナリズム)、本田康二郎氏(現象学)、八代嘉美氏(幹細胞生物学・科学技術社会論)の6名が学会での議論に参加した。

また、公開ワークショップ後のディスカッションには、小山田和仁氏(科学技術政策)、小山虎氏(哲学)、成原慧氏(情報法)、野澤聡氏(科学史)、平川秀幸氏(科学技術社会論)の5名にもご参加いただいた。

本ワークショップは、国立情報学研究所の平成27年度公募型共同研究と科学研究費助成事業挑戦的萌芽研究「人工知能の規範・倫理・制度に関する対話基盤と価値観の創出」の助成を受けて開催した。

### 目次

- ※ 第5回ワークショップ報告
- ※ 第6回ワークショップ報告
- ＊ ワークショップ概要
- ＊ アカデミアの役割とは
- ＊ 共通言語基盤を作りあげる
- ＊ 段階を踏んで考えていく
- ＊ 開発慎重論と積極論
- ＊ 規制と責任の在り方
- ＊ データの処理と公共性
- ＊ まとめと今後の展望

まず、ワークショップ企画者4名が話題提供を行った。最初にオーガナイザーの江間有沙（科学技術社会論）が趣旨説明を行い、続いて服部宏充氏（マルチエージェントシステム）がAIRグループの目的を概説し、異分野間協働研究は「つながる」だけではなく相互学習して「つなげていく」努力が必要であり、その方法を模索するべく現在アンケート調査やインタビュー調査を実施していると説明した。また、人工知能学会倫理委員会の活動を紹介した。

次に秋谷直矩氏（エスノメソドロジー）は、社会学者は問題が発生したあとの整理・分析は得意とするが、情報系研究者からは「君たちは批判するばかりで解決してくれない」と言われてしまうジレンマを紹介した。これに対し、社会学者としては10年、20年の長期的な視点から持続可能な異分野間連携を可能にする組織設計を考えたり、1980年代の第2次人工知能研究ブームなど過去プロジェクトの検証を行うことで、今後の協働研究に資する知見を提供していきたいと述べた。

最後に大澤博隆氏（ヒューマン・エージェント・インタラクション）は、人工知能は人間の価値判断や意思決定に干渉しうる技術であるため、価値関数決定のためには人工知能研究者が人文知を手に入れるか、人文・社会学者にエンジニアへ近づいてもらうなどの歩み寄りが必要であると述べた。その上で「人工知能の浸透する社会」のデザインをどのようにしたいかを、多くの人と議論していくことが重要であると最後に結んだ。

続いて外部からの話題提供者として矢島章夫氏（国立研究開発法人科学技術振興機構社会技術研究開発センター）が人工知能を中心とする情報技術の発展とリスクに取り組む組織としてFuture of Humanity Institute (FHI:英国)とFuture of Life Institute (FLI:米国)を紹介した<sup>1)</sup>。特にFLIのファンドは、倫理的観点のプロジェクトやポスドクなどの若手研究者のプロジェクトに予算が一定割合配分されていると指摘した。

次に高橋恒一氏（理化学研究所生命システム研究センター/全脳アーキテクチャイニシアティブ理事・副代表）が、汎用人工知能（自律的に学習する人工知能）の研究拠点が国内外で増えてきたことと、理化学研究所での人工知能研究拠点について紹介をした。また、これから出現する技術について考える枠組みとしてELSI (Ethical, Legal and Social Issues)に代わり、HELPS (Humanity, Economics, Law, Politics and Society) という概念を提唱した。

高橋氏とともに2015年からAI社会研究会を主宰している井上智洋氏（マクロ経済学）は、労働経済学者David Autorらの実証研究<sup>2)</sup>やオックスフォード大学のシミュレーション結果<sup>3)</sup>を参照して、1990年代に労働市場で二極化が発生していることや、人工知能が雇用に与える影響を論じた。一方で、人工知能を上手に利用することで経済成長が急速に進む可能性をシミュレーション結果から示した。

次に株式会社富士通研究所の中尾悠里氏（科学技術社会論）が、技術の設計や実装にあたってプライバシーやセキュリティなどの問題が生じないように、企業は顧客との対話を何度も行って製品開発をしていることを説明し、その対話プロセスや手法を紹介した。その小さな対話の延長上にAIRなど学術的な研究グループが掲げている「社会の価値」があるのではないかとの意見が述べられた。

最後に、井上悠輔氏（公衆衛生学・医療倫理）はIBMワトソンがガン診断に導入されようとする中<sup>4)</sup>、本来であるならば医療業務を「支援」する道具でもある技術が、いつの間にか業務自体の在り方を変えるかもしれないこと、またそれによって生じるだろう問題を指摘した。例えば医者だけでなく一般の人も人工知能を診断に使用できるようになると、医療者間、医者と患者間での判断・責任の所在を考える必要性がでてくる。また、人工知能を使いこなすことが医学教育で重要となるかもしれないなどの変化の可能性を指摘した。

話題提供者の発表後、会場の参加者とともに議論を行った。次ページから、議論の論点を項目別に再構成したものを示す。

## 注

- 1) 西下佳代, 茅明子, 矢島彰夫, 奥和田久美, 2015 「人工知能やロボットの社会的影響に関する先行的研究動向」, 第30回研究・技術計画学会予稿論文集.
- 2) David H. Autor et al., 2006. "The Polarization of the U.S. Labor Market." *American Economic Review*, 96(2): 189-194.
- 3) Frey C. Benedict & Osborne A. Michael, 2013. "The future of employment; how susceptible are jobs to computerization?" Oxford University Engineering Sciences Department.
- 4) 「東大医科研が「Watson Genomic Analytics」を使ったがん研究を開始IBMプレスリリース」  
<http://www-06.ibm.com/jp/press/2015/07/3001.html>

## アカデミアの役割とは

AIRは2つの目的を掲げている。1つは「政府による干渉や産業による利益誘導に左右されない異分野間の対話・交流を促すための媒体や基盤をボトムアップで構築すること」である。これは政府や産業を排除して議論を行うということではない。本ワークショップも政策や企業の視点から話題提供をいただいた。もう1つの目的は「対話を通して、人工知能の目指すべき共通アジェンダや社会の未来ビジョンを設計し、技術開発・実装時の新設計基準や規範・倫理・制度に関する価値観を提案すること」である。これはAIRのみが行っているわけではない。人工知能がインストールされた製品やサービスの制度設計が企業などですでに開始している現在、本研究グループがある種固定されたメンバー、かつ大学研究者が中心となっている意味について改めて考えねばならないだろう。大学には、様々な分野の知と価値が蓄積されている。多様な分野の研究者との共通言語基盤を構築し、協働する体制を築くことで、新しい知や人材を生み出すことがアカデミックな研究グループに期待されていることではないだろうか。

## 共通言語基盤を上げる

本ワークショップでは「(汎用)人工知能」「人文知」「論理的」などの単語が意味するものの確認作業が度々行われた。また単語だけではなく説明の仕方の違和感も議論された。例えば人工知能は「学習する」「見る」など人間の所作をメタファー(比喩)として用いる。しかし、現在の特化型の人工知能は最適解の発見が目的であり、あくまで「学習しているように見えている」ものである。それに人間のメタファーを用いると混乱を招き、人工知能脅威論に飛びがちである。一方で、人間のような知能を作りたい人工知能研究者にとって、メタファーを用いることは研究のモチベーションを高める。また逆にメタファーを使っていくことで概念が拡張されることもある。人工知能のイメージや概念が固定していないからこそ、単語レベルからの合意形成が重要であることが確認された。



## 段階を踏んで考えていく

「汎用人工知能に、何ができないのかを教えてください」という質問に対し、研究者によって回答は様々であるだろうという留保があった上で、「何年までに」という制限がなければ、およそ人間にできることは実現できるだろうが「できる」と「社会が許容する」ことは別物であるとの議論があった。

人工知能と社会の相互作用は、段階を踏んで社会との関係性の中で考えていく必要がある。例えばIBMワトソンによる医療診断が導入されることで、すぐ医者代わりに人工知能が判断を下すことにはならないだろう。まずは医療診断の道具として医者が人工知能を使うことから始まる。そこでの相互作用によって、次の段階の人工知能の開発の方向性や医者や患者の役割、価値観が変容する可能性もある。そのために現在の価値基準のまま50年、100年後の技術の在り方を語るのではなく、5年、10年などの段階を踏んで考えていくことも重要である。

## 開発慎重論と積極論

先進的な技術開発をする時に、事前警戒原則(precautionary principle)という考え方がある。例えば遺伝子組み換え作物など、問題が起きたら影響は甚大であると想定された場合には、影響や因果関係が十分に示されていなくても規制措置を適用できる。人工知能の開発も、雇用や生活への影響が大きいと想定されるため、開発は慎重にすべきではないかとの意見もあった。

しかし情報技術は他分野ほど巨額の資金を必要としないことに加え、グローバル化している現在、技術は黒船のように押し寄せている。人工知能を持つ、あるいは使いこなせる人や国とそうでない人や国との間でデバインド(格差)や雇用の破壊が起きる可能性は否定できない。

ほかにもIBMワトソンの医療診断など、一企業が市場を独占することの問題もある。例えば日本に症例の多い胃がんや肝がんのデータを提供した上で、システムを購入させられるというデータの収奪が起きるのではないかとの懸念も指摘された。これは人工知能に限らず先進的な技術一般について言えることであり、実際に生命科学分野では、装置がほぼアメリカ製になっていることが指摘された。また独占は経済的な問題を引き起こすだけではなく、システム内部がブラックボックス化することで、再現性確認が困難になるという研究倫理上の問題も含む。

## 規制と責任の在り方

事前警戒より強力なものとして法規制がある。規制には事前規制と事後規制がある。前者は運転免許証のように使用許可や資格を定めたもので、後者は問題発生後の責任や被害の特定方法を定めたものである。高リスク分野では事前規制が導入され、介入が強くなる。事後規制は問題発生後の対応という反応の遅さはあるが、自由を確保するという点は重要である。

また規制に至る前から、どのようなリスクがあるのかを整理し議論することが重要である。特に人工知能に価値判断や意思決定を委ねるようになった場合の責任の所在が議論された。現在は人工知能の価値関数は人間が与えている。そのため、どのような価値や決定を是とするかの議論が重要である。しかし、価値や決定に関する社会的合意が取れさえすれば、問題の解決は難しくないかもしれない。例えば、自動走行車の事故時の責任の所在は、損害賠償責任等の法的な責任をどのように分配するかという問題であり、保険によって解決できる部分もある。

一方、人工知能自らが価値判断や意思決定を学習から導き出せるようになった場合は、問題がより複雑となる。例えば医療診断の材料となる症例を自律的に学習していくこと、あるいは人間の子供のように倫理観をコミュニティから学習していく人工知能が実現した場合、その意思決定や価値判断に誰が責任を持てるだろうか。人工知能の製造者だろうか、使用者だろうか、あるいは学習した人工知能本体だろうか。

## データの処理と公共性

このような議論はSF的になるので、もう少し身近な例に引き寄せて考えてみたい。例えば現在、地震速報などのニュースはすでに人工知能が作成している。統計的に処理できる問題は、ディープラーニング技術の進展によってさらなる飛躍が期待できる。

しかし統計的に処理できないこと、文脈に依存することに関しての学習は困難を伴うだろう。特に玉石混交のデータセットを扱う場合、何をジャンク（ごみ）データと判断するかは価値の問題を含む。その判断精度を向上させるためには、「質の良い」大量のデータを用意する必要がある。逆に悪意あるデータを意図的に読み取らせば、悪影響をもたらす機械を作り出しかねない。倫理ジョークとして「人間は必ずしも理想的な行為者ではないので、人間から倫理を学習してはならない」という意見もあった。

また、読み込むデータセットが異なれば、人工知能の下す判断や決定にも個性や個体差が出てくるだろう。例えばマイノリティの価値判断や決定を代弁する人工知能が市場を席巻した場合、マイノリティがマジョリティになるというジレンマが生じる。そのような場合、価値判断や意思決定の公共性や民主主義とは何かということを再考する必要が出てくるだろう。これは何も新しい話ではない。現在でも、我々の見ているニュースや得ている情報は編集され時にカスタマイズ化されているのだから。

## まとめと今後の展望

本ワークショップは2時間で9人が話題提供するというかなり密なスケジュールであった。企画者側の話題を改めて振り返ると、自分たちの分野を持ちながら他者の分野に歩み寄るための方法論や組織の在り方を模索している内容となっていた。結成から約1年以上たった現在、互いの思考や目的が徐々に融合してきている。しかし、そのような時期であるからこそ、今後の方向性や問題設定は慎重になる必要がある。

今回のワークショップでは外部の方5名に話題提供いただいた。政策・汎用人工知能・経済・企業・医療に関する議論の多くは、今までのワークショップで話題に上りつつも深堀できていなかった論点でもある。今回、専門家の話を伺うことによって、人工知能と社会に関する議論は幅広いものであるということが再確認された。

議論しつくしたというには到底時間が足りないが、ワークショップは交流の場でもある。本ワークショップでの議論をきっかけとして、これから5年、10年先の未来を見据えた異分野間対話の活動を今後も続けていきたい。



AIR Webサイト

URL: <http://web4ais.wpblog.jp/>

AIR NEWS LETTER vol.2 (2)

編集: ARISA EMA