OS-5

人と AI が織りなす新たなエコシステム

A New Ecosystem Interwoven between Human and AIs

中川　裕志　理化学研究所革新知能統合研究センター
Hiroshi Nakagawa　Center for Advanced Intelligence Project, RIKEN.
iroshi.nakagawa@riken.jp, https://sites.google.com/site/nakagawa3/home

山川　宏　株式会社ドワンゴ ドワンゴ人工知能研究所
Hiroshi Yamakawa　Dwango Artificial Intelligence Laboratory, Dwango Co., Ltd.
hiroshi_yamakawa@dwango.co.jp, http://ailab.dwango.co.jp/

Keywords: data-ecosystem, AI ethics, AGI, accountability, responsibility, cooperation, well-being.

1. OS の発足とその狙い

本オーガナイズドセッション（OS）「人と AI が織りなす新たなエコシステム」は，2017 年の人工知能学会全国大会（JSAI 2017）に向けて新たに企画・申請し採択された新設 OS である．

企画および実施は，著者らのほかに，浅田 稔（大阪大学），井上智洋（駒沢大学），今井倫太（慶應義塾大学），江間有沙（東京大学），金井良太（(株) アラヤ），高橋恒一（理化学研究所），萩田紀博（(株) 国際電気通信基礎技術研究所, 以下 ATR），堀川優紀子（ATR），松原繁夫（京都大学）らの 11 名により進められた．さらに本学会汎用人工知能研究会（Sig-AGI）の協賛を得ている．

1·1　本 OS 発足の背景

本 OS を設置した背景は，人工知能（AI）が急速に進展し，人々の生活の中に深く浸透し，その便益のみならずリスクへの危惧も高まっていることがある．例えば，データ処理によるバイアス，職業への影響面が指摘されている．また中長期的には，貨幣経済，資本主義，法制度，倫理観，労働の価値，民主主義などといった社会的枠組みも影響をもたらし得る．

背景となる高度 AI の進展による影響についての大局的分析は本 OS 内でも行われている．著者の一人山川は「速度耐性をもつ社会の実現にむけた基礎的考察」（発表番号：1F3-OS-5b-01）を発表した．ここでは，さまざまな変化が加速し，いずれ AI には自律性を導入することが不可避となるシナリオを示した．つまり，個別に導入する AI が道具的で制御可能であったとしても，それらがネットワーク化すると人手による制御に限界が生ずる．例えば，その変化の速度が人の認知能力を超えたり，変化が多すぎて対応できなくなったりする．このため導入された自律的 AI が具体的手段を講じる大元となる最上位の目標もしくは価値観をいかに設定するかが最重要となる．なお，関連記事は本誌 2019 年 3 月号（Vol. 34, No. 2）の特集「道徳判断の自動化をめぐる問題」内に掲載予定である．

高橋恒一は「将来の機械知性に関するシナリオと分岐点」（発表番号：1F3-OS-5b-03）を発表した．ここでは長期的な機械知性の行き着く形として，上限シナリオ，生態系シナリオ，多極シナリオ，シングルトンシナリオなどを定義し，それが，物理的な側面を含むさまざまな計算技術的な制約により分岐することを議論している．未来の分岐は人類の未来にとって，AI がいかに独占されるか，そうではなく共有物になり得るのかといった可能性にも関わる重要な問題である．この問題に対して技術的背景をもった議論を行ったことは極めて斬新である．なお，関連記事は，本誌今号（2018 年 11 月号, Vol. 33, No. 6, pp. 867-872）のレクチャーシリーズ「シンギュラリティと AI」において掲載される．

1·2　本 OS 発足の狙い

人の能力が AI で拡張されたとしても，人と AI が織りなすシステムの全体は大自然のごとく複雑広大であり，個人としても人類全体としても，それを完全には理解し把握できない．

こうした AI が進展する背景を踏まえれば，人々が速すぎる変化から守られつつあまねく恩恵を得られ，なおかつ持続可能な形をつくり出すためのエコシステムを新たに構築することは急務であると思われる．こうした変化を踏まえ，さらなる変化に備えるために，私達がなすべきことは数多いが，例えば，AI は不具合やハッキングを避けながら高度なロバスト性をもたせる必要があ

る．AIと人間の対話の在り方はいかにあるべきか．さらに技術進展にキャッチアップしたリスク管理や法制度整備を行うためにマルチエージェントシミュレーションなどを用いる必要も生ずるだろう．AI自身がもつべき価値観（公平性・公正性を含む）や，AIが占めるべき法的および倫理的な地位（人格）や責任についての研究も必要である．

今後多様な懸念事項に対処していくためには，多くの人々の協力による多面的なアプローチが必要である．学術的に見れば，計算機技術のみならず，倫理，哲学，経済，社会，心理，法律，歴史，政治，社会などのさまざまな分野の専門家の知識や意見を取り入れていくことが有効であろう．さらに，当学会としては，日本の文化や価値観を生かしながら世界的なコラボレーションに向けて貢献できるよう，技術的な議論を深める場をつくることは有意義であり，本OSがこうした活動の連携に貢献できれば幸いである．

2. 本オーガナイズドセッションの発表

このオーガナイズドセッションは6月5日午後に行われた．冒頭に中川による趣旨説明が行われ，その後に，一般発表8件，招待講演1件が行われた．（a）学生が自分の将来の仕事とAIの関係をどう見ているか(1件)，(b)社会が人工知能とどう向き合うかの分析とモデル化と国によるモラル獲得の相違（2件），(c) AIの行動の社会的責任の在り方（2件），(d) 自動運転車の実装実験として駐車場での駐車などの実験結果（1件），(e) 高度AIの進展とその影響の大局的分析（2件）などが発表され，最後に戸田山和久氏の招待講演が行われた．

（a）の発表は主として文科系の大学生が，AI技術が彼ら自身の目指す職業にどう関わると考えているかを調査したものであったが，AI技術を否定的に捉える意見，肯定的に関わろうとする意見，哲学的問題として思考停止する意見などが抽出された．まとめると，大学学部の低学年ではAIについて曖昧な知識しかもっておらず，彼らの本性的感覚が表れているようである．これはAI技術に十分なイメージをもたない一般的な人々にも通じるような傾向と考えられ興味深い．と同時に，AIが社会的に重要な役割を果たすような時代においては，十分な知的対応策が見えていないことを示している．よって，AI技術者としては一般社会へのAI技術や概念の普及に力を入れる必要があることが見て取れる．

図1 会場の様子

すでに（e）高度AIの進展とその影響の大局的分析に関わる2件の発表については触れているので，以下では（b）のAIの社会的責任を分析した2件の論文および招待講演についての報告を行う．

2·1 AI と 責 任

AIの行動に人間はどのくらい責任をとらせることができるかという論点は本オーガナイズドセッション「人とAIが織りなす新たなエコシステム」の本質的テーマの一つである．この論点に関して三宅智仁らの心理的側面からの分析論文「人工物との共同作業における心の知覚と責任帰属の関係」と赤坂亮太の法律的側面からの提言論文「不法行為法におけるAIの法的人格に関する検討」は異なる切り口から重要な知見を与えてくれる．以下でこれらについてやや詳しく報告する．

○人工物との共同作業における心の知覚と責任帰属の関係（発表番号：1F2-OS-5a-2）三宅智仁ほか

今後，増加が予想される自律型ロボットと人間の共同作業において，作業に失敗し，損害が生じたとき人間はその責任分担をどの程度ロボットに求めるかを以下のようなゲームの場で実験した結果について述べているのが本論文である．

実験で用いたゲームは，参加者と相手が二つの選択肢（「たくさんほしい」と「相手に譲る」）のどちらかを選択し，その組合せによって金銭的報酬が与えられることを10回繰り返すゲームである．参加者と相手の選択肢に組合せによって以下のような正負の報酬がおのおのに与えられる．

表1 報酬表

		参加者の選択	
		たくさん欲しい	相手に譲る
相手の選択	たくさん欲しい	参加者：−10円	参加者：＋20円
		相手：−10円	相手：＋100円
	相手に譲る	参加者：＋100円	参加者：0円
		相手：＋20円	相手：0円

実験における参加者と相手の組合せは次の3種類である．

1）人間の被験者（Aさん）に対して，相手も人間
2）人間の被験者（Aさん）に対して，相手は有体物のロボット
3）人間の被験者（Aさん）に対して，相手はソフトウェア

この 3 種類の組合せのおのおのに対して，上記の報酬表によって対戦ゲームを行う．ゲーム後，累積した報酬額が予測される期待値より低くなった場合に，人間の A さんが敗因を相手の責任にする度合いを A さんに質問して測定した．また，これと並行して，A さんが相手側の記憶力，道徳的，自制力，コミュニケーション力，計画性，思考力，感情的をどのように見積もっているかも質問して調査した．

この実験評価の結果，

(1) A さんが，相手側が記憶力，自制力，計画性，思考力があるとみなす度合いが高いほど，相手側の責任を軽減して感じることがわかった．しかし，直観的には人間はロボットよりもこれらの 4 項目で大きな力をもっているので，相手が人間より責任能力が高いはずである．ところが，

(2) 実験では相手が人間の場合のほうが相手の責任が大きいとみなす割合が多いという結果が出た．

この (1)，(2) は矛盾する結果である．そこで，相手側の記憶力，道徳的などの項目に加え，意識をもつこと，苦痛を感じること，個性をもつこと，などの心の知覚項目についてゲームの前後での変化量を被験者 A さんに質問して調べてみた．すると，道徳的，苦痛を感じることという心の知覚の度合いがゲーム前後で大きく下落すると，相手の責任を問う傾向が高くなることが有意にいえた．つまり，相手が人間の場合は，A さんは自分が負けた場合に相手の道徳的，苦痛を感じることの度合いの下落傾向を強く感じる．つまり，人間である相手に自制がないとみなす度合いが大きいと相手側の人間の責任を強く感じる．一方，ロボットの場合は，これらの下落傾向が小さい．つまり感情的に行動していないとみなしているため，心の知覚項目はゲーム前後で変化しないため，その責任を問わないということになる．

なお，相手側への責任帰属に関しては，ロボットのほうがソフトウェアより人間に近いとみなす傾向も抽出された．言い換えれば，無体物あるいは抽象的存在であるソフトウェアよりも有体物で具体的存在であるロボットを人間に近いとみなし，人間のような意識をもつことを期待する傾向がある．このことから，責任帰属の大きさは，人間＞ロボット＞ソフトウェア　という評価実験結果となる．

以上が三宅らの論文の主張であるが，現実世界においてはすでに AI による金融トレーダーや AI による保険査定などソフトウェアが多大な責任をもつ社会的ないしビジネス的行為を行っており，ロボットや自動運転車などの有体物はこれらに比べてまだ大きな社会的，経済的影響をもつといえない．したがって，この論文の結果は，社会での AI ソフトウェアの利活用の現実に即するような拡大を狙って展開することが望ましい．その展開によって，実世界で影響力のある知見が獲得できると期待できるようになる．

○不法行為法における AI の法的人格に関する検討（発表番号：1F2-OS-5a-2）赤坂亮太

半自律的ないし自律的な AI や AI を搭載したロボットの社会実装は現在すでに進行し，その責任に関わる問題を起こしている．例えば，自動運転車は自律的 AI とみなせるが，試験走行で事故を起こしている．また，ダ・ヴィンチ外科手術システムも AI 搭載ロボットであると考えられるが，手術ミスを起こしている．こういった事故はアメリカにおいては裁判になり判例も出始めている [平野 18, パガロ 18].

このような事態において AI ないし AI 搭載のロボット（以下，「AI・ロボット」と表記する）の事故における責任の取り方を考察しておくことは喫緊の社会的課題である．赤坂はこの問題を AI・ロボットに人格を与えるかどうかという切り口から出発して議論している．この論文では，AI・ロボットが引き起こした不法行為の責任の所在を議論している．不法行為の責任に関する法律である不法行為法とは，故意または過失によって生じた損害に対する被害者の救済として損害補填を行うことに加えて，不法行為の抑止効果も狙うものである．

AI・ロボットに財産権も含む法的人格を与えれば損害補填は可能である．しかし，不法行為の抑止効果の狙いは，1）AI・ロボットは自然人ではないので，感情的な反省は期待できない，2）AI・ロボットに関わる保険制度が充実すれば，補填をそれに頼ればよい，のような考え方が生まれる，という理由で法的人格を与えても効果は期待できないとしている．さらに従来の機械類より AI・ロボットの行動が予測しにくい，つまり予見可能性が低くその動作内容がわからないブラックボックスになっているという実情を考えると，自賠責保険のような制度，あるいは製造物責任のような考え方が有力となってくる．しかし，自賠責保険は保険料が必要であり，製造物責任は明らかに AI・ロボットの開発側に負担が大きく，どちらも開発者を委縮させるだろう．

そこで，赤坂は責任を社会全体で分担し開発者の責任を免除する無過失補償制度を提言している．このような制度はすでに「ニュージーランド事故補償法」という先例もある．さらに法律家から見た日本の法体系において無過失補償制度の位置付けを紹介し，AI に法的人格を与えるよりも優れた考え方であることを論証している．

ここまでが赤坂論文での提言だが，赤坂は AI・ロボットが結果の予見段階において十分に緊張し回避義務を行ったかどうかを法律家が問題にすると述べている．さらに，AI のブラックボックス化がこの回避義務を困難化しているという理由で予見可能性に基づく過失責任を退け，無過失保障制度を提言している．しかし，AI 技術においては透明性，説明可能性，説明責任などの議論が盛んに行われており，技術的にもう少し様子を見たいところである．例えば，著者らの私見ではあるが，対象 AI を十分多様な環境で試験使用して，その動作を外部

から観測することは可能である．さらにこの観測結果として得られた膨大なデータを機械学習技術によって分析することもできるので，対象AIの動作を高い精度で予測するような別の観測・予測AIを開発する方向はあり得るだろう．この観測・予測AIが予見可能性にかなり広範囲に資する可能性も否定しきれないところである．

2·2 招待講演：戸田山和久（名古屋大学）

○我々はなぜ未来社会を構想することが下手なのか，じゃあどうすればよいのか

戸田山氏は哲学者である．そこでこの招待講演では，哲学者の視点から以下に述べる課題についてお話をうかがった．

講演の課題：幸せな未来社会に向かうためにどのように科学・技術を生かすかを考えることが大事だということをほとんど自明なこととして受け入れたうえで，AIの未来について実り豊かな議論をしたい．だが，そのためには，幸せな未来という概念を構築することが難しい．この難しさの原因にまでさかのぼって再考する．

以下では戸田山氏の講演の流れに従って，内容を順次説明していく．

現在のAIに技術開発は主に与えられた仕事を効率化するタイプの機械を発明するという目標に向かって動いている．これは，一見荒唐無稽に見えても具体的な目的が与えられた場合にそれを実現するドラえもんのポケットからのツール取出しに似ているのでドラえもん症候群と呼ぶ．しかし，与えられた目的を効率的に実現しようと研究，開発，実用化をしているだけでは幸せにはなれない．つまり，すべてが想定内の物事の実現なので，感動は起きない．通常，AIの研究者や開発者は決められた目的実現の効率化を意図して研究・開発を行っているが，それが人間にとって幸せなものかどうかを考えることはほとんどしないし，またそのように考えることはむしろ不得意に見える．不得意な理由としては，これまでの工学の教育が与えられた目的の効率化を達成するようなスキルの習得だったこと，未来の我々が何を望むかを想像することが難しいことなどがあげられる．

そこで発想を変え，「AIやロボット，すなわち機械自体は与えられた仕事を効率良く実行できなくても，何かはわからないけれど，今まで存在しなかった新しいことができる可能性を与えるものであれば，人間に幸せをもたらす」と幸せを定義してみる．例えば，初期のAppleのマッキントッシュはほとんど何もできなかったけれど，新しいことを始める糸口つまり可能性を与えてくれた．そこからいろいろなアイディアが生まれ，世の中が変わっていったといえるだろう．

このような幸せの定義は以下のような疑問に答えることができる．つまり，偉大な芸術家は一生，創作活動に苦しんでいる．しかし，果たして不幸だろうかと問われれば，不幸ではないと本人も言うだろう．なぜなら，芸術家は生きている限り，常に新しい創作の可能性が自身の内側にあると思って仕事や思考をしているからだ．このような将来の可能性が断たれることが死であり，それは不幸なことだろう．芸術家に限らず，人間は未来に可能性があると幸せであり，死はその可能性を完全に消し去るため不幸だということになる．

以上の考察から得られるレッスンは，以下のようになる．AI研究者・開発者にとって，幸せな未来社会を想像するのは通常は難しいが，それにもかかわらず，幸せな未来社会をもたらす技術を構想したいなら，以下のような発想を取り入れることが有力である．

1）便利なものをつくろうとするのをやめて，人間の可能性を拡張するツールを社会制度の設計と並行して発想せよ．

2）幸せな社会（ユートピア）よりディストピアのほうが想像しやすいなら，まずディストピアを想像し，それをユートピアに転化する方法を考えよ．

この2）の項目は少し説明が必要であろう．ユートピアは意外に想像しにくい．著者もそのように感じている．実際，美術館に行っても，戦争，死，地獄などの絵は非常に多く傑作に富むが，ユートピアや天国の絵は少なく傑作もあまりない．つまり，人間はユートピアよりディストピアを想像するほうが得意である．ならば，まず最初は全力でディストピアを想像し，その対極としてユートピアすなわち幸せな社会を捉え，ディストピアをユートピアに変えていく方法を研究・開発の目標にすることを推奨しているわけである．

戸田山氏は，ここで，当初の課題に対する一応の結論を得たわけだが，今後の課題は，上記1），2）のアイディアを教育プログラムに落とし込むことであるとして締めくくっている．

3. 今後に向けて

AIの社会への影響についての議論は，世界的に見て2016年頃から活性化し，2017年にはおおむね主な論点が出そろい，ある程度具体的な議論が進み始めた．すでに，2018年においてはさまざまな組織がその立場に応じて論点をピックアップした形でAIと社会に関わる何

図2 大懇親の様子

らかの表明を行うようになってきている．例えば，マイクロソフトは「Future Computed：人工知能とその社会における役割」という文書を公開したし，Google は開発原則を発表している．こうした中で，アカデミアとしての私達が，どのような形で世の中に貢献することが可能であるかは大きな問いである．

一方で全国大会（JSAI 2018）では，本 OS を開催した 6 月 5 日の夜に「AI と社会」大懇親会を開催した．ここでは当学会の倫理委員会のほかに，三つの OS「人工知能と倫理」,「複雑化社会における意思決定・合意形成のための AI 技術」,「自律・創発・汎用 AI アーキテクチャ」のメンバーらとともに 60 人規模の懇親会を開催した．ここで当学会における AI 技術の研究だけでなく，社会との境界領域をカバーする研究会の設立についての話題も出た．

◇ 参 考 文 献 ◇

[平野 17] 平野 晋：ロボット法，弘文堂（2017）
[パガロ 18] ウゴ・パガロ 著，新保史生 監訳，松尾剛行，工藤郁子，赤坂亮太 訳：ロボット法，4.2 節，pp. 99-107，勁草書房（2018）

著 者 紹 介

中川 裕志（正会員）

1953 年生まれ．1975 年東京大学工学部卒業．1980 年同大学院工学系研究科修了．工学博士．1980 年より 1999 年まで横浜国立大学工学部勤務．1999 年より 2018 年まで東京大学情報基盤センター教授．2018 年より理化学研究所革新知能統合研究センター・グループディレクター．人工知能と社会制度，人工知能倫理，プライバシー保護技術などの研究に従事．情報処理学会，電子情報通信学会，言語処理学会，Association of Computational Linguistics，IEEE などの各会員．

山川 宏（正会員）

1989 年東京大学大学院理学系研究科物理学専攻修士課程修了，1992 年同大学院工学系研究科電子専攻博士課程修了．工学博士．同年，株式会社富士通研究所入社後，センサフュージョン，RWC プロジェクトに参加．現在，株式会社ドワンゴドワンゴ人工知能研究所所長，NPO 法人全脳アーキテクチャ・イニシアティブ代表，電気通信大学大学院情報システム学研究科客員教授，玉川大学脳科学研究所特別研究員，産業技術総合研究所人工知能研究センター客員研究員，慶應義塾大学 SFC 研究所上席所員，理化学研究所革新知能統合研究センター客員研究員，東京大学医学部客員研究員．人工知能，特に，概念獲得と操作，汎用人工知能，認知アーキテクチャ，ニューロコンピューティング，意見集約技術などが専門．神経回路学会，電子情報通信学会などの各会員．共著書に，「パターン認識と機械学習」（シュプリンガー・ジャパン，2008），「人工知能とは」（近代科学社，2016），「宗教と生命　激動する世界と宗教」（角川書店，2018）などがある．