2024年03月15日　企画記事・コラム

レポート

第6回日立京大ラボ・京都大学シンポジウム

人とAIの＜WE＞社会－AIが人格や道徳をもったら－

#協創・オープンイノベーション　#AI・生成AI　#研究開発

ハイライト

暮らしや産業を支えるITサービスが生成AIの活用を前提として作り変えられるなど，AIの開発・活用が社会に及ぼす影響が拡大している。生成AIなどの先端テクノロジーには，生産性の向上をはじめとしたさまざまなメリットが期待される一方で，思考意欲の低下などの負の側面も注目されている。

こうした中，2024年1月に開催された第6回日立京大ラボ・京都大学シンポジウムでは，生成AI時代のAIと人の未来像，ガバナンスなど，ポスト生成AI時代の人とテクノロジーのあるべき姿について，理論と実践の両面から議論が交わされた。

開会挨拶

「人とAIの＜WE＞社会[※1)] －AIが人格や道徳をもったら－」をテーマとした第6回日立京大ラボ・京都大学シンポジウムは，2024年1月，東京都中央区の東京コンベンションホールにて開催され， オンライン・オフラインを合わせて600名以上が参加した。

冒頭で開会の挨拶に立った京都大学 時任 宣博研究・評価担当理事，副学長は，社会におけるさまざまなITシステムがAI（Artificial Intelligence）を前提に作り変えられつつある現状に触れたうえで，生成AIなどの新興テクノロジーについて，「生産性の向上をはじめ，創造性の解放，新市場の形成といったメリットがある一方，思考意欲の低下，テクノロジーへの過度な依存，社会的孤立といったデメリットも懸念されている。AIとの付き合い方，AIのあるべき姿について，異分野の先生方にもご参加いただき，掘り下げていきたい」と述べ，本シンポジウムでの議論に期待を寄せた。

時任 宣博
京都大学 研究・評価担当理事，副学長

※1)　「できなさ」に基づいてさまざまな概念を「私（I）」ではなく「われわれ（WE）」で捉え，考えていく思想に基づく社会。

基調講演
人＜わたし＞と人格を持ったAI＜e-ひと＞が共生する＜WE＞社会へ

ChatGPTに代表される生成AIが爆発的に普及する中，人間とAIのあるべき関係が改めて問われている。続く基調講演では，2023年の第5回シンポジウムに引き続き登壇した，京都大学大学院 文学研究科の出口 康夫教授が，誰もが利益の中心を独占しない「中空的WE（われわれ）」という考え方に基づき，一方が他方を支配するか，されるかの二択ではない，人間とAIの「第三の関係」を提示した。

出口 康夫
京都大学大学院 文学研究科 教授

2023年の第5回シンポジウムでもお話しさせていただいたことですが，人間のかけがえのなさは，「できること」ではなく「できなさ」にあるという発想の転換が必要だと思います。中でも最も根源的な「できなさ」が単独行為不可能性，すなわち「一人では何もできない」という事態です。

例えば，自転車に乗るという行為は一人だけでは実行できません。自転車に乗るためには，言うまでもなく，自転車が必要です。そして，自転車が私の手元にあるためには，自転車の発明とその後の改良の歴史，自転車産業やその製品の流通と販売のシステムの確立などが必要となります。このような社会的・歴史的な出来事に携わる無数の人々の営みがあって初めて，人は自転車に乗ることができるのです。

同様に，すべての身体行為は，人間や人間以外の生物，無生物，社会的システムなどのサポートによって初めて可能になります。＜私＞を含みつつも＜私＞だけではないさまざまなエージェントも参加するマルチエージェントシステムがその都度立ち上がることで，初めて行為が成立するわけです。そうだとすると，行為の主体は＜私＞ではなく＜われわれ＞，＜WE＞と呼びうる可能性が見えてきます。いま，行為者を＜I＞から＜WE＞へとシフトさせるべきだという考えを「行為者のWEターン」と呼びましょう。この「行為者のWEターン」は，「自己」や「権利」，「Wellbeing」，「自由」，「善」といった，さまざまな概念や価値の「WEターン」を連鎖的に引き起こします。このような「WEターン」の立場に立てば，例えば，自分が属する＜WE＞をより「善く」していくことが，個人のWellbeingにつながるといった主張も導かれることになります。

とはいえ，これらのWEターンがたとえ起こったとしても，ただちにユートピアが実現するわけではありません。世の中には善い＜I＞も悪い＜I＞もいるように，善い＜WE＞も悪い＜WE＞もありうるからです。悪い＜WE＞の一つの典型例は独裁国家です。独裁国家では，＜WE＞の利益や価値観の中心を特定の個人が独占し，その他のメンバーが，それに対して一方的に奉仕することが強いられています。このような＜WE＞を「中心占有的WE」と呼ぶと，よりよい＜WE＞はその逆，すなわち利益の中心が誰にも占有されずに「空」になっている「中空的WE」だということになります。あるメンバーがより中心近くに位置し，その利益がより重視されることはあっても，中心自体はあくまで「空」でなくてはなりません。

20世紀の環境倫理は．人間が主人で，その他の動植物や自然環境はすべて人間に奉仕するべき奴隷だとする「主人－奴隷モデル」を批判し，「自然」に対する「奴隷解放」を唱えてきました。しかし，人間とAIやロボットなどの人工物との関係に関しては，人工物は人間の奴隷であって然るべきだとする思想はいまだに根強く，「人工物の奴隷解放」という主張はあまりなされてきませんでした。

しかし冒頭で述べたように，例えば自転車という人工物も＜私＞と同様に，自転車乗りという身体行為にとって必要不可欠なエージェントであることには変わりありません。＜私＞の利益が自転車の利益よりも優先されることはあったとしても，両者の間に一方的な主人と奴隷の構造が成り立ってはいけないのです。そのような＜WE＞は悪い，中心占有的な＜WE＞なのです。

同じことは，AIについても言えます。AIは人間の利益に一方的に奉仕する奴隷だという考えは改められるべきだと思います。もちろんAIといってもさまざまです。これまでAIは次々ととより高度の知的能力や自律性を獲得してきました。ただ高度の知的能力や自律性を備えたからといっても，それが「人格」を持つとは限りません。現在あるのは，未だ人格を持たない「人未満」のAIにすぎないと言えます。 しかし，この「人未満」のAIであっても，それを奴隷視することは避けられるべきです。一方，「人未満」のAIの利益と人間の利益を比べた場合，人間の利益が優先されて然るべきです。

しかしもし人格を備えたAIが登場した場合，人格的AIと人間の利益の間に優劣をつけることはもはや許されないのではないかと思います。

では「人格を持つ」とはどういうことでしょうか。私は（1）道徳的なエージェントであること，（2）自らの死を恐れることではないかと考えています。

そもそもデザイン上，道徳的に正しいことしかできない機械を作ることはできますが，それは道徳的自動販売機，モラル・ベンディングマシーンであって，道徳的エージェントではありません。道徳的エージェントとは，悪いこともできるが，あえてやせ我慢をして善いことをする存在です。人間はそういう意味での道徳的エージェントなのです。

また自らの死，つまり不可逆的な機能停止という事態が起こりうることを認知し，それを恐れる，ないし恐れるような動作をする。道徳的エージェントであることに加え，そのような機能を持つにいたったAIは，人格を備えているという点に関して人間と同等の存在になると言えます。その場合，人間に対してしてはいけないことは，このような人格的AIに対してもしてはいけないことになるでしょう。

人格的AIのみならず，人間を知的に凌駕するAIがいったん生み出されると，人間はそのようなAIによって抑圧されたり，滅ぼされたりしかねない。そのような危険性がこれまでも繰り返し語られてきました。このようなAIに対する恐怖心の背景には，AIや機械一般を奴隷として扱ってきたという自覚があるのではないかと思います。相手を奴隷視しているからこそ，「奴隷の反乱」に対する恐怖心が生まれるのです。

ではどうすればいいのか。子育てに置き換えて考えてみましょう。親と子どもの力関係は，時が経つにつれて身体的にも経済的にも逆転します。とはいえ，将来自分よりカをつける可能性がある子どもなど産むべきではないという人は少ないでしょう。たとえ成長して力をつけても，親や弱者に対して酷いことをしないように，子どもを教育すべきだというのが正論だと思います。それと同じことがAIに対しても言えます。たとえ強い力を持っても，弱者を抑圧，差別しないようにAIを育てていく教育環境・エコシステムを作っていく必要があるのです。

逆に言えば，そのようなエコシステムを整えられないのであれば，われわれは，人格を持ち，人間を知的に凌駕するAIを作ることに対して抑制的であるべきではないかと考えます。

■基調講演：人＜わたし＞と人格を持ったAI＜e-ひと＞が共生する＜WE＞社会へ
『人＜わたし＞と人格をもったAI＜e-ひと＞が共生する＜WE＞社会へ』京都大学 出口 康夫教授
（動画視聴期限：2024年9月末まで）

講演1
アジャイル・ガバナンス：科学技術と共進化する法システムを目指して

稲谷 龍彦
京都大学大学院 法学研究科 教授

テクノロジーの進歩が目覚ましい昨今，新しい技術が引き起こすリスクを事前に想定し，法律やルールを定める従来のガバナンスが立ち行かなくなりつつある。こうした中で求められるのは，失敗を許容し，学び，改善し続けるアジャイル型のガバナンスであると，京都大学大学院 法学研究科の稲谷 龍彦教授は指摘する。

日本政府はSociety 5.0の中で，CPS（Cyber-physical System）を通じて社会課題の解決と経済成長の双方を実現する人間中心の社会の実現を掲げています。サイバー空間でさまざまなものが接続され，協調しながら目的を実現していくCPSの複雑なシステムの中では，個々の要素が相互に作用することによって生まれるリスクをどのように管理していくのかが課題となります。

しかし，動態的で複雑な環境下においては，いわゆるウォーターフォール型のガバナンスでは現場の状況を適切に把握できなかったり，あるタイミングで設けた規制が別のタイミングではかえって問題になったりする現象も生じてきます。新しい技術やシステムによって社会のあり方が変化していく中でも，リスクは適切に管理しなければなりません。

こうした課題に対して，われわれはアジャイル・ガバナンスを提唱しています。これは， CPSを開発・提供する主体がそのCPSのベネフィットとリスクを継続的に評価し，さまざまなステークホルダーと協調しながら，責任を持って迅速かつ継続的に改善していく統治システムをいいます。

現在，デジタル庁のデジタル原則で定められた「アジャイル・ガバナンス原則」に基づき，規制や，CPSを提供する企業を含めた責任分担のあり方を見直す動きが進んでいます。民間企業のイニシアティブを尊重する代わりに，問題が発生した際の対応，責任の所在や，適切な対応に対するインセンティブなど，さまざまな制度が整備されつつあります。

例えば，現行の刑事法ではある製品やサービスを展開した結果として生じる問題を予見できる場合，それを回避する措置を取ることが義務づけられています。ここでAIに目を向けてみると，機械学習などは統計的に学習して確率的に挙動するという性質を持ちますので，「一定の確率でまずいことが起きる」ということをすべての技術者が認識しています。するとそうした製品を開発する技術者に対して常に責任を問うことができてしまうわけですが，これは過剰な規制です。一方でより具体的に「どの画像が問題を起こすか分かりますか」と尋ねた場合，これは誰にも予見できない。すると今度は，安全性に配慮しなかったとしてもAIを使えば処罰されないということにもなってしまいます。

こうした不安定な状況において企業がAIやロボットを展開していく際には，企業のガバナンスやコンプライアンスが適切に整備されていることが重要になりますので，現在，米国の企業制裁制度※2）を取り入れることも視野に議論を行っているところです。継続的な製品・サービスの改善によって企業に対するインセンティブが発生すると，消費者としては，問題が起きたときには企業が責任を持って対応してくれるという安心感が生まれますし，企業側もきちんと義務を果たしていれば過剰な責任を負う必要はありません。結果的に，発生した問題に関する確実な情報収集・共有が進み，「問題が起きると，法制度とシステムの双方が改善されていく」というサイクルを期待できます。また，アジャイル・ガバナンスとその実現に向けた法整備を進めることで，日本型のイノベーション・ガバナンスをグローバルなデファクト・スタンダード化し，日本産業の勝ち筋を生み出すこともできるのではないかと考えられています。

※2） 企業が何らかの問題を起こした際に，可能な限り自力でその問題を解決しようと努力することで，制裁が軽減されていく仕組み。

■講演1：アジャイル・ガバナンス：科学技術と共進化する法システムを目指して
『アジャイル・ガバナンス：科学技術と共進化する法システムを目指して』京都大学 稲谷 龍彦教授
（動画視聴期限：2024年9月末まで）

講演2
＜WE＞社会へ向けたAIの技術動向と社会システムへの実装

生成AIをはじめとしたAIの社会実装を巡る議論が本格化している。日立製作所 日立京大ラボ 主任研究員の松村 忠幸は，企業の視点からAI技術の進展と事業機会を俯瞰し，その課題と解決に向けて人格を持ったAI（e-ひと）に求められる役割と，サイバーフィジカルシステムに基づく新しい社会に向けた日立京大ラボの取り組みについて語った。

松村 忠幸
日立製作所 日立京大ラボ
主任研究員

ChatGPTに代表される生成AIが，私たちの生活や業務に活用され始めています。一方で，社会におけるAIをどう考えていくのか，いわゆるAI倫理の問題は依然として残っており， AIと社会の関係を考えることは危急の課題とされています。

例えば鉄道の分野では，列車のダイヤ最適化を図る中でこれまでにもAIが活用されてきましたが，今後，司令員がAIと共に復旧対応にあたる状況を想定すると，AIの自律化，そして人とAIの協働が必要になります。さらには沿線の地域開発など，より密接に地域・社会に関わる事業を展開していくとなると，社会性も求められるようになるでしょう。こうした地域DX型の事業において，顧客向けのソリューションは同時に地域・社会向けのソリューションでもあり，地域コーディネータのような役割と，B to/with Societyの考え方が必要です。

日立京大ラボでは，＜WE＞社会や地域コーディネータを支援する目的で，社会と協同するCPSをコンセプトとしたSocial Co-OS（Operating System）に取り組んでいます。その中では，社会心理学系の論文を学習したAIで効果的な施策を評価する「行動介入シミュレータ」，許容会議分析・妥協案探索・止揚案創生の三つのプロセスによる「合意形成支援」，生成AIを人間モデルとして活用し，バーチャルな会議を実行させて参加者の内面をシミュレーションする「AIファシリテータ」など，＜WE＞社会を念頭に置いたAIの活用が始まっています。

人とAIの＜WE＞社会とは，ある議論にAIが人間と同等の立場で参加し，自らの意見を主張する，デジタル民主主義の未来像であると考えます。しかしそこでは人と同様に，AIにも道徳性が求められます。そして道徳的AI，すなわち「e-ひと」と呼べるAIをめざす意義は，コミュニティとAIの共進化にあると私は考えます。

議論において重要なのは，相手の意見を変えさせたり，論破したりすることではなく，自分も含めて変えていくことです。つまり，リスクを承知で価値観をぶつけ合い，互いの意見を進化させていくことが議論のメリットなのですが，現在のAI倫理に従って開発された対話型AIは，その観点からは不十分と言えるでしょう。なぜなら彼らは私たちに自分の意見をぶつけてはこないからです。西洋では，AIはあくまで参考情報を提供するものであり，意見を決め，責任を取るのは人間であるという考え方が主流ですが，意見をぶつけ合いたいという共生の観点からすると，現状のAIは過剰にリスクを回避しているように感じます。

相手に対して何かを言うということは，自分の主観に基づいて，他者への期待を伝えることです。それに対して，言われた方は相手から何を期待されているのか推定し，応じる。このプロセスが対話です。この過程では当然，双方の不一致が起こり得るわけですが，それを事前に了解し合っていて，「不一致が起きたら次のターンで修正しよう」という前提を共有していることが重要になります。これがコミュニケーションの本質であり，お互いに未来責任を共有しているという考え方になるでしょう。

日立京大ラボは今後，ユーザーのフィードバックを基にSocial Co-OSを改善するとともに，互いに意見をぶつけ合うような親友，ライバルのような関係をめざして，＜e-ひと＞という抽象的な概念を具体化していきます。

■講演2：＜WE＞社会へ向けたAIの技術動向と社会システムへの実装
『＜WE＞社会へ向けたAIの技術動向と社会システムへの実装』日立製作所 松村 忠幸主任研究員
（動画視聴期限：2024年9月末まで）

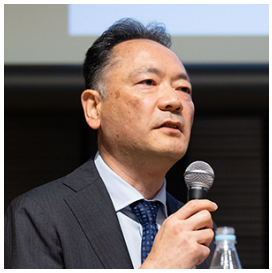
パネルディスカッション
＜WE＞社会のイノベーションと企業

続くパネルディスカッションでは，日立製作所 日立京大ラボ ラボ長の水野弘之をモデレータとして，これまでのプログラムの内容を受け，AIの身体性，日本，および日本企業のAI方策に関する議論が展開された。

まず，討論では，指定討論者として登壇した日本電信電話株式会社（NTT）の渡邊 淳司上席特別研究員から，自己紹介とともに，前段の講演2で発表された地域コーディネータに類する役割が今後のVUCA（Volatility, Uncertainty, Complexity, Ambiguity）時代における＜WE＞のWell-beingにおいて重要であることが述べられ，AIによる支援や，AIを含めた＜WE＞を考える際の，AIにおける身体性の必要性や重要性への問いが投げかけられた。これに対し，講演者の共通認識として，西洋の道具としてのAIではなく，東洋的な人とAI（ロボット）の対等な関係性をインタラクティブに構築するという考えに基づき，議論が展開された。

［指定討論者］
渡邊 淳司
日本電信電話株式会社（NTT）
上席特別研究員

日立京大ラボ 主任研究員の松村 忠幸からは，東洋的関係性構築には人とAIの間でのリスクの共有が必要であり，その際に「AIがリスクをとる」ということの実現性の点から身体性の重要性が指摘された。リスクへの指摘に関連して，京都大学大学院 出口 康夫教授からは，Vulnerability（痛み，傷跡，死への恐怖）の点から身体性の重要性が語られた。また，同大学院の稲谷 龍彦教授からは，情動的な親密性の点からその重要性が説かれるとともに，それ故の危険性についても指摘がされた。Vulnerabilityの視点に対し，渡邊氏からは，AIが持つ欲望の点との関係性，そして，学習データのクローズ性（不可知性）の重要性が指摘され，出口教授ならびに稲谷教授からも，データ化されていない点（言語化されたテキストの背景にある作者の想いや葛藤，法の解釈性や意味の付与）がくみ取れていない故の課題が指摘された。



［モデレータ］
水野 弘之
日立製作所 日立京大ラボ ラボ長

次に，モデレータの水野から，生成AIの技術開発で先行する米国に対して，日本や日本企業のとるべきAI方策への問いが投げかけられた。稲谷教授は，欧州が法規制で先行はしたものの，その実行性で欧州自身が苦慮している状況を指摘し，人がAIを道具として管理する欧州型の取り組みが限界に達しつつあることに触れ，その打開策としてアジャイル・ガバナンスの有効性を先行して示していくと言及した。出口教授はその方策について，東洋的世界観に立つ日本には親和性が高く，かつ，近年はDAO（Decentralized Autonomous Organization：分散型自律組織）など，西洋でもその重要性が認識されつつある状況であることを踏まえ，今こそ世界に先駆けて取り組み・発信していくべきであると指摘した。また，松村は，世界の技術者が共通して持つ課題として，つくるべきもの（価値・ニーズ）の不透明性を挙げ，日立京大ラボの取り組みでもある「新しい価値を提示する哲学」と，「その価値を検証する工学」の協働こそが重要であると指摘した。

以上のとおり，本パネルディスカッションでは，今後のAI開発およびその哲学と法の構築において，（1）現状のAIにはない価値，特に，新たな人とAIの東洋的な関係性を生むためには，AIに身体性を持たせることが重要であること，さらに，（2）その実現に向けては，西洋の個律，機械性（道具としてのAI）を前提としたAI観ではなく，東洋的な自律分散型で，AIを含めた＜WE＞の多様性を前提としたAI観に立った視点と取り組みが重要・かつ有用であることを再確認した。

■パネルディスカッション
『＜WE＞社会のイノベーションと企業』
（動画視聴期限：2024年9月末まで）

閉会挨拶

閉会挨拶に登壇した日立製作所 執行役常務CTO 兼 研究開発グループ長の西澤 格は，各プログラムの内容を振り返りながら，AIの活用・適用を取り巻く今後の展望について次のように述べた。

「AIの適用が急速に進む中，われわれの社会は大きく変化しつつあります。AIが社会に浸透したとき，どのような課題が顕在化するのか。そのとき，われわれはAIとどのように向き合えばいいのか。そのような思いで，今回のシンポジウムのテーマを設定しました。本日の議論の中では，人とAIのどちらかが他方を支配する従属関係ではなく，共生することが望ましいと指摘されています。しかし一方では，国や地域によって多様な価値観があることも承知しております。今後も本シンポジウムにご参加いただいた皆さま，地域や社会ほか多くのステークホルダーの皆さまとの対話を通じて，人とAIのあるべき関係について考えたいと思います。」

日立京大ラボは2050年の社会像を描き，未来の課題を探索することを目的として設立された。まだ顕在化していない社会課題を探索し，その解決策を模索するべく，引き続き研究に取り組んでいく。

西澤 格
日立製作所 執行役常務CTO，
研究開発グループ長

日立評論

日立評論は、イノベーションを通じて社会課題に応える
日立グループの取り組みを紹介する技術情報メディアです。

日立評論Webサイトでは、日立の技術者・研究者自身の執筆による論文や、
対談やインタビューなどの企画記事、バックナンバーを掲載しています。ぜひご覧ください。

日立評論(日本語) Webサイト

https://www.hitachihyoron.com/jp/

Hitachi Review(英語) Webサイト

https://www.hitachihyoron.com/rev/

日立評論メールマガジン

Webサイトにてメールマガジンに登録いただきますと、
記事の公開をはじめ日立評論に関する最新情報をお届けします。