特　集「AI and Society」

今考えるべき問題と社会へのインパクト（懸念と期待）

Near-Term Issues and Impacts

青山　俊之　筑波大学
Toshiyuki Aoyama　University of Tsukuba.
turetiru@gmail.com, http://www.turetiru.com/

Keywords: artificial intelligence, AI safety, industrial revolution, human and AI partnership, beneficial AI.

1. は　じ　め　に

本稿はAI and Societyの2日目の特別セッション「今考えるべき人工知能への社会へのインパクト」，セッション1「今考えるべき問題と社会へのインパクト」におけるケンブリッジ大学・CFIのエグゼクティブ・ディレクターであるスティーブン・ケイブ氏，慶應義塾大学法学部教授の大屋雄裕氏，北京大学の燕京奨学生であるダニット・ガル氏の三つの講演内容を要約して報告したのち，質疑応答も紹介する．

2. AIの倫理・自律性の問題

ケイブ：AIは素晴らしいものか，恐ろしいものか．ある一つのものが同時に二つの特徴をもつことはあり得ます．AIを素晴らしいものにしたいのなら，解決すべき問題があります．AIが及ぼすマイナスの影響への対応を考えていけば，問題を避けることができるかもしれません．

短い時間ではありますが，ガバナンスの問題についてお話しします．今後数年でAIを良い方向に開発していくために考えなければならない問題です．AIとは，計算機がアルゴリズムを走らせ，大量のデータを扱うものです．AIの倫理は計算機の倫理・技術分野の倫理と，かなり重なります．現在のAI革命は機械学習のおかげで，そのマシンのエンジンになっているのはデータです．

現在，データ倫理にとどまらず，もっと新しい倫理指標が出てきています．それはこの2日間で議論されている「知能」と「自律性」です．AIが知能と自律性をもつようになると，子供が大人へと成長していく中で倫理的な立場も変わってくるように，AIも責任が求められるようになるのです．

まず古典的なデータの倫理問題があります．プライバシーはデータガバナンスに関する重要な問題ですが，それ以外のデータ関連の問題は自律性や知能といったAIに関する問題と関連しています．例えば，バイアスの問題です．データセットは限定的で，特定の方法で特定のグループから特定の期間に集められたものになります．そのデータセットに偏見があれば偏見があるままの出力が返ってくるでしょう．例えば，検索エンジンで「経営者」と検索したときに，男性の画像が多く出てきます．社会が偏りをつくってきたからです．データ倫理は自律性によってより複雑になります．アルゴリズムでインプットのバイアスを排除できるかもしれないですし，逆により強化されてしまうかもしれません．自律性が高まることで，データの問題がより複雑になります．

では，自律性とは何でしょうか？　AIが魅力的なのは，私達の代わりにタスクをこなしてくれるからです．AIはよりスピードが早く，安く，時には人間よりもタスクを上手にこなします．自動運転でも1秒ごとに「運転を続けますか？」と聞かれるくらいなら自分で運転したほうが良い．そうではなく，AIに任せられるのが自律性です．

自律性の問題とデータの問題で，重なっている問題の一つが透明性です．現在の機械学習システムは決定木など人間の概念に則ってプログラミングされておらず，AI自身がデータの扱い方をプログラミングしていきます．そのため人間の概念に翻訳するのは難しく，人間にとってわかりにくいシステムとなってきています．自律的学習の問題だけでなく，自律的行動の段階で問題は悪化します．例えば，住宅ローンをシステムは許可するのか，あるいは歩行者が道路に出てこようとするときに自動車は止まるのか迂回するのか．そういった決定をAIが自律的に行うと，誰が責任を取るのかが非常に曖昧になります．人間の尊厳を損ねるこ

とになるため，生死に関わるような決定はマシンに委ねてはならないという意見もあります．例えば自律型兵器など，対象を選んで殺すという決定を，自律的に行うことを認めてはならないという意見があります．戦場だけでなく，誰が医療のケアを受けるべきなのか，保釈されるべきなのか，そういった人生に大きな影響を及ぼす決定を機械が行っても私達は安心していられるでしょうか．

また，価値観の整合性をどう取るかという問題もあります．意思決定や社会における価値の一貫性をどう保持するか，AGI（汎用人工知能），ASI（人工超知能）を構築する際にも重要です．私達よりも賢い機械をつくるなら，私達の価値観と一致させなければなりません．

自動化で仕事が失われるのは新しいことではありませんが，真面目に受け止めなければならない問題です．産業革命によって自動化が起こり，その結果ファシズムや共産主義の台頭をもたらしました．二つの世界大戦が起こったことにも一部起因するといわれるほどです．現在は，産業革命時よりももっと早いスピードで開発が進み，グローバルに機械が使われるようになっています．そうなると，もっと破壊的な影響を及ぼす可能性もあります．全く新しい問題もあります．第一次産業革命は人間の筋肉の力にとって代わるものをつくることが行われましたが，現在は私達よりももっと賢い機械をつくっています．そのため人間に残された領域はないのではないのかという倫理的な問題があります．やることがなくなった人間はどうするのか？福祉の提供や再訓練を施さなくてはならないでしょう．また尊厳ある生活，仕事で定義されない生活の議論が必要となります．仕事から解放されたいが，社会に必要とされたいというパラドックスにも取り組んでいかなければいけません．

知能と自律性をもつことにより，AIはより多くのタスクをこなすことができるようになりますが，その結果，人間がAIに依存してしまう危険もあります．人間は経験から学びますが，機械が代替することで人間は経験の機会を失います．自動運転時に，突然人間が運転しなくてはならなくなった場合，今の私達はできるでしょう．しかし運転を学ばなくなった子供達が，自分で運転しなくてはならなくなったらどうなるのでしょうか．あるいは，医者が診断をAIに任せていたら，医者自身で診断するスキルを失ってしまわないでしょうか．

私達は批判的な思考能力を失ってしまうのかもしれません．機械のほうが優れていると思うなら，機械に任せてしまう．AlphaGoを開発したデミス・ハサビス氏を，最初は懐疑的に見る人が多かったそうです．しかし，AlphaGoが勝ち始めると批判ではなく，好奇心をもつ人が増えてきました．AlphaGoは今や世界王者となって，疑問視されることはなくなりました．こうした中，批判する思考能力を失ってしまうのは，囲碁のようなゲームならよいかもしれませんが，医療や輸送システムの現場においては重要な問題かもしれません．

AIの開発を否定したいわけではないことを強調しておきます．上記の問題は，AIを最も良い方向に開発していくためには考えなければならない問題です．違った国，文化，コミュニティでは，それぞれ懸念や解釈に違いがあり，互いの意見に耳を傾けることが必要だと思います．

現在は歴史的な瞬間で，知性の革命だと後になって歴史家は言うでしょう．責任ある持続可能性のある形で，将来的にもメリットがある形で開発していくチャンスでもあります．

3.「責任の裂け目」を避けるための枠組み

大屋：私は法哲学者で，法学部で仕事をしています．マイケル・サンデルで有名になったトロリー問題を考えてみましょう．路面電車が制御不能で暴走しており，このままだと右に行って五人の作業員をひいてしまいます．ポイントを左に切り換えると代わりに通行人を一人ひくことになります．ポイントのすぐそばにいるあなたはどうすべきかというものです．では，このポイントがAIによって制御されている場合，どのようにプログラミングすべきでしょうか．

この問題について適切な答えを導くのは非常に難しいことを強調し，人間の尊厳を守ることができない限り新しい技術を開発すべきでないと強調する哲学者が多くいます．しかし注目すべきなのは，人間が操作する場合でもこの問題は解けていないということです．責任の分配という問題に注目しましょう．車の事故では，何らかの過失があれば運転手に責任を課すことができます．責任は生じた損害を社会的に分配する制度だと考えることができ，結果を左右できた人に責任を割り当てているわけです．このシステムを，過失責任と呼びます．

しかしレベル4の自動運転車の事故の場合，誰に責任があるのかは非常に難しい問題です．このような状況を，私は責任の裂け目と呼んでいます．車内に座っている人は制御できないため，責任がありません．自動運転車は自己学習プロセスを経ています．メーカは学習の条件や枠組みについて決めることができますが，結果として選択される行為すべてを予測はできません．それはちょうど，子供の育て方を選ぶことはできても具体的な行為は予測できず，責任を負えないのと同じことです．

この問題の解決策として，無過失責任を導入することが考えられます．日本では例えば原子力事故に関し

て特別法で認められており，福島第一原子力発電所の事故においては東京電力の過失を証明することなく賠償が得られています．強制保険という対応を考えることもできます．日本で自動車を運転するためには，一定の保険に加入しなくてはなりません．これも特別法によって対応することができるでしょう．

しかし問題は，そのような法制度をつくるためには国会に至る複雑な手続きが必要だということです．政治家，官僚，そして一般の人達の同意を得る必要があります．それまで問題を放置しておけば，事故に対する適切な賠償は行えないことになるでしょう．既存の制度が新しい技術に解決できなくなると，一般の人達は驚き，いらだち，パニックを起こすかもしれません．これは，技術にとっても社会にとっても長期的には良いことではないでしょう．安定的に健全な技術の発展を守っていくためには，何らかの形の解決策を適切に形成していかなくてはなりません．

一つの取組みを紹介します．日本では，総務省のAIネットワーク社会推進会議が，国際的な議論のためのAI開発ガイドラインを発表しています．大学関係者，経営者，エンジニアなどの複数のステークホルダが作成に関与しており，私も委員の一人です．これは非拘束的なソフトローです．

基本原則にはAI技術によって人間中心の社会を実現することを掲げ，国際的なステークホルダとの間でベストプラクティスを共有することを目指しました．連携，透明性，制御可能性，安全など九つの原則で構成されており，すでに文書としてWebで公開されています．重要な点はこれがまだドラフトであること，技術的な発展に基づいて改定されるものだということです．さらに利活用ガイドラインをそれぞれの分野で作成する計画もありますが，いずれにせよ国境を超えた議論が必要です．今回のシンポジウムは，このような問題について考える良い機会になると考えております．

4. 中国のAI計画

ガル：中国のAIの計画について紹介します．今，AIに関する政策で先導しているのは日本です．内閣府の懇談会，総務省のAI開発ガイドライン，人工知能未来社会経済戦略本部の動きや人工知能学会倫理委員会の倫理指針などを見ていくと，日本がリードしていることがはっきりとわかります．米英もレポートを出していますが，まだ道は長いと読み取れます．

一方，中国は大胆な展開を三つ打ち出しています．第一に中国は自分達がこの分野でリーダーになると宣言しました．第二に政策改革として基礎研究を強化し，AI分野の下支えをすると述べています．第三に世界のAI開発のリーダーになると宣言しました．中国は，2020年までに他国に追いつき，2025年に大規模なブレークスルーをAIで行い，2030年にはリーダーになる計画を立てています．そのための三つの柱があります．1本目の柱として，新世代AI理論技術の開発，2本目の柱としてAI技術の標準・基準が中国企業と適合するように基幹エンタープライズをつくること．これはすでに行っています．3本目の柱として倫理規範，政策・規制をつくることを掲げています．

1本目の柱としての新世代AI理論技術の開発競争はデータ保護と新たなサイバーセキュリティ法のもとで行われています．中国では7億のネチズンが大量のデータを生み出しており，彼らが競争ではなく協力することで問題解決につながります．この計画に多くの資金が投資されています．第2の柱としての標準・基準には政治的な戦いがあります．多くの国は自分達のやっていることに基づいた標準をつくりたい，他の国に使ってほしいと考えています．ところが標準があれば，グローバル競争を進めることができます．第3の柱として文化特異的な倫理規制を技術に組み込むことが考えられます．技術の開発時には，ある特定の文化が織り込まれているので，輸出先の国の文化を変えるほどの影響をもちます．そのため，倫理的な規範と能動的な規制が必要であることに理解があります

第2段階のAIで突破口を開くために，次世代AI技術の開発を掲げています．そのために基礎研究に着目していますが，これに対して機微な発見があった場合の透明性について懸念されています．例えば，素晴らしい研究が共有されなかったらどうなるでしょうか．基礎研究に関しては，特にこの点が懸念されています．例えば，中国のAI業界が世界市場を独占したらどうなるでしょうか．あるいは，欧米と建設的で対等な交流を築くまでに発展したらどうなるでしょうか．欧米の会社ではなく中国が主導する機会も増えてくるかもしれません．欧米と中国で安全や管理の定義が異なった場合，それがぜい弱性を招かないでしょうか．安全は誰が決めるのか，それを誰が実践するのか，これによって技術開発も変わってきます．しかし協力すれば，効率的な安全性，管理メカニズムができてリスクを抑えることができるかもしれません．この点については現在考えなければなりません．

第3段階に至ったら中国はリーダーになると宣言しています．次世代AI技術と理論を主導し，基礎研究によってリーダーシップを取り，世界トップの業界をもつのだといっています．包括的なAIに関する法律や倫理規範ができるだろうという想定のもと実効性のある規制が期待されています．

一方で二つの懸念があります．一つは技術の分断化です．ある技術が開発されたとしても国や文化が違うと使い方が全く違うのかもしれません．そのため協力をして文化や価値の多様性を尊重していくことが重要

です．二つ目の懸念として，AIの力が強くなることによって，政治でいうところの相互確証破壊（MAD：一方が先制的に使うと結果的に相互に破壊し合うことを確証する核戦略の概念）のようなものになるかもしれません．競争の方法を間違えれば，このような危険性があります．

一方でコラボレーションすることによって実効性のある監視が可能にもなります．AI開発を先導する国同士が手を組んで，技術がみんなのためになるための協力が必要です．現在，さまざまな国がAI開発を主導しようと戦略を発表しています．それに対して公的な監督やコラボレーションのメカニズムは，まだ道半ばであり，政策立案者が能動的に手を組む必要があります．AI開発していく中で各国とも能動的に十分に考え抜いた規制が必要になります．どの国も競争環境に置かれている状況で，規制を話し合うことは効率が悪いのは確かですが，社会のことも考えないといけません．技術は国境を越えます．日本で良い技術が開発されたからといって日本にとどまるわけではありません．技術開発をするすべての者は近隣に影響を及ぼすわけです．その責任を取る必要があります．選ぶのは私達です．

野心的で大胆なことを言っていると思われるかもしれません．でも，できる可能性はあります．長期的に思われるかもしれませんが，実は短期的な話なのです．現在，競争はすでに始まっています．このままだとゼロサムゲームで終わってしまいます．最後に，私の生き方を変えたアドバイスを皆さんに送ります．

「変えることができないものは受け入れなさい．自分が受け入れることができないものは変えなさい．我々が選択肢をもっています．我々が変える力をもっています．ゼロかイチだけではありません．」

5. 質 疑 応 答

質問1：ケイブ氏はAIの自律とおっしゃっていましたが，人間の自律もAIの影響を受けるかもしれません．どのようにお考えになっていますか？

ケイブ：人間が技術に依存してしまうリスクもありますが，AI技術が人間の自律を高め，特に厳しい暮らしを強いられている人々へのエンパワーメント，より多くの人達が自分で決断する手助けになるとよいと考えます．

質問2：ガル氏は文化特異的な規制技術が埋め込まれるとお話しされましたが，中国の興味深い例を教えてください．

ガル：ソーシャルパートナーロボット研究は欧米より東アジアが中心的で，子供のお守りだけではなく，兄妹や友人，教育者としての役割を果たすロボットが提案されています．中国では一人っ子政策を背景にこういったロボットが非常に広く受け入れられています．二人以上の子を産む親のプレッシャーを緩和することもあるかと思います．

質問3：ガル氏の「AIはゼロサムゲームである」とは具体的にどういうことでしょうか．また，どうすればそれをストップできるとお考えでしょうか？

ガル：国々は競争の中で成長し，戦略的に国の資産をつくるためにはゼロサムゲームにならざるを得ないと考えています．それに陥らないようにするためには，次世代のAI開発を共に行うことが必要でしょう．安全性を確保するためにも，競争にルールを入れるべきです．

質問4：全員に質問です．短期的なAI開発と規制のバランスについてどう思われますか？

ケイブ：規制と開発の間には緊張があります．政府がエンジニアの合意なしに管理しようとしてもうまくいきません．またAI技術は一般的にグローバルな性格をもつため，ある国で発明をされたとしても，他国やネット上で使われます．管理のため厳しく規制をしようとしても効きません．政府は研究の自立も尊重する必要があります．そのため日本の総務省で規制ではなく原則としてソフトローを考えているというのが私の考えです．

大屋：開発への規制を懸念していた企業も，今では私達の議論に加わってくれます．産業界は議論を避けたいのではなく，確実な答を求めているのだと思います．ただ，問題の答えを得るには時期尚早です．ですから原則をつくってモニタしていく．そして何か深刻な問題が起こったら政府がすぐに対処できるようにするアプローチが必要だと思います．

ガル：日本のSociety 5.0で，政府はテクノロジーを開発することによって社会を改善していく政策ビジョンを出しています．このビジョン実現と規制は一貫しているのか．それを考えるためにも開発者は，技術が誰に対してどのような影響を及ぼすのかを伝えることが重要です．

質問5：中国の掲げるシナリオは反証不可能であると疑っています．また，米国の法的な分野では理論的に機械の自律を扱うものがないため問題になっています．システムがわからないとき，法律専門家は管理ができるのでしょうか？

ガル：北京に住んでいなければ，私も同じ疑問をもったでしょう．ただ，第1の2021年の目標はかなり現実的で，第2の2025年の目標も一定程度は達成すると思います．2030年の目標についてはまだ何ともいえません．中国ブレインプロジェクト，中国ブレイン2.0という二大プロジェクトは期間限定のプロジェクトでしたが，これも予定どおり進むと思います．

ケイブ：すでにパワフルな自律的エージェントが使われています．私達のケンブリッジ大学のセンターでは，

法人概念，社会の法律認識，重要な決定を下すエージェントの扱いを考えています．法律は保守的なものであるべきですが，追いつく必要があります．

大屋：ロボットの責任問題を，ローマ法における奴隷の扱いによって解決する試みもあります．法的な道具は蓄積されており，暫定的な解決策を与えることは十分に可能ですが，それが問題の本質を捉え損なう危険性も意識する必要があります．

質問６：どうすれば競争ではなくコラボレーションを促すことができるのでしょうか？　誰がそのような立場にあるのでしょうか？

ガル：しっかりした基準・標準をつくって皆が遵守すれば，最終的にはAIは今のWi-Fiみたいになるかもしれません．多くの企業が協力も競争もしていますが，少数の強い企業があると市場の不均衡を生みます．政策立案者がビジョンをもち，産業界は利益をあげようとしますが，究極的にビジョンをもって実用化するのは私達です．

質問７：ケイブ氏は広く議論をしましょうとおっしゃいましたが，2016年にはBrexitやトランプ大統領の誕生があり，人々の意見は分断されています．

ケイブ：Brexitとトランプ大統領台頭ではソーシャルメディアを筆頭とした技術が人々の意見を分断しているといわれており，それにも対応する必要があります．多様な人々の対話は確かに難しいですが，女性の声やグローバルな観点からいえば地球の南側の人々の声を聞くことも大事だと思います．

質問8:全員に質問です．午前中のセッションでベン・ゲーツェル氏が分散型システムのお話をされました．これは，ゼロサムゲームから離れる一つの方法かと思います．政策の中で分散システムは合理的に動くでしょうか．

ケイブ：ゼロサムゲーム緩和の実現可能性はあると思いますし，そう願っております．AIの経済を分散するのであれば，誰しもが向かっていきたいことだと思いますし，非常に価値のあることだと思います．

大屋：独占と健全な競争とでは後者のほうが有益だというのが歴史の教訓であり，だから市場経済が支持されてきました．分散型システムが健全な競争のプラットフォームとして，皆が理解して合意できるルールのもとでフェアな舞台になるのであれば，ほとんどの関係者がそのプラットフォームを選ぶでしょう．

質問９：AIを使った検閲システムを計画している国があると聞いています．政策立案過程に学者がほとんど関与できない国もあります．我々AIの研究者はこういった状況に対してどうすべきだと思いますか？

ケイブ：不安は理解できますが，原子力兵器とは異なり，検閲に関わるAI技術は，市民が政府に対して使うこともできます．NGOから政府，あるいは行政から別の行政のチェックを行うこともあり得ます．これもある種の分散化システムであり，競争，あるいは協力を健全に維持するものだと思います．研究者が悪意に勝てると約束はできませんが，そうしていくしかないと思います．

文責者あとがき

「今考えるべき」とあるように，シンギュラリティが起きている，起きるという議論以前にどのようにAIが開発され，政策としても進められ，現代社会に浸透しているのかが今回の議論で垣間見えた．ケイブ氏が取り上げた自律性の問題に，大屋氏が示した法的責任の問題と日本における取組み，そして最後にガル氏が紹介した中国のAI計画．企業における開発と法的関係，社会的立ち位置の関係だけでなく，国策としてもAI開発と社会整備を主導するうえでの競争が進められている．本シンポジウムは研究者に限らず多様な人が集い，交流と議論が交わされた．しかし，技術が分断を起こしている側面も否めない．先行きの不透明な，理論的な枠組みとしての「正しさ」を安易に主張できない局面において，いかに社会包摂を進めることができるのかが大きな課題なのではないかと思う．その際には，良い悪いという二元論ではなく，多面的な分析と判断が必要だろう．そうした価値の判断をするうえでの議論はやはり人がせざるを得ない．当然，議論はAI技術がもたらす影響についての議論が大半であったが，「人」がいかに技術を使うことができるのかに対する議論にも今後，より注目していきたいと感じた．

2018年2月2日　受理

著　者　紹　介

青山　俊之
1992年生まれ．筑波大学社会・国際学群国際総合学類4年．