

「人工知能と人間社会に関する懇談会」

報告書

平成29年3月24日

人工知能と人間社会に関する懇談会

目次

Executive Summary	----- 1
第1章 はじめに	----- 3
1.1 人工知能技術が社会にもたらすインパクト	
1.2 人工知能技術がもたらす期待と不安	
1.3 人工知能技術はこれまでの技術と何が違うのか	
1.4 懇談会の目的	
第2章 人工知能と人間社会に関する検討動向	----- 6
2.1 世界の動向	
2.2 日本の動向	
参考 人工知能は人間を超えるか？	----- 8
第3章 懇談会におけるアプローチ	----- 10
3.1 分野の選択	
3.2 検討する観点	
3.3 共通する論点の整理へ向けて	
第4章 人工知能技術と人間社会について検討すべき論点	----- 13
4.1 倫理的論点	
4.2 法的論点	
4.3 経済的論点	
4.4 教育的論点	
4.5 社会的論点	
4.6 研究開発的論点	
第5章 おわりに	----- 19

Executive Summary

現在、人の知的活動（認知、思考、推論、それらに基づく行為等）を部分的に代替しうる人工知能技術は、ビッグデータと機械学習によって急速な発展をしており、自動運転車や医療診断支援、対話エージェントなどへの実装も進んでいる。人工知能技術は、日本政府が目指すSociety 5.0の重要な基盤技術であり、少子高齢化がもたらす労働力不足などの社会課題の解決や誰もが自分の能力を発揮して活躍できる社会づくりに貢献し、社会に多大な便益をもたらすことが期待されている。ただし、人工知能技術は、その開発速度が速く利用者からはその動作が見えにくい技術であることから、知らぬ間に普及し高度化し社会の在り方に根本的影響を与える可能性もあり、健全な利用のためにその影響を検討する必要がある。

国内外において、人工知能技術が人間社会に及ぼす影響についての倫理的、法的、社会的問題（Ethical, Legal, and Social Issues: ELSI）が注目されており、研究者・開発者のみならず、様々な人が人工知能技術のもたらす良い影響と懸念に関心を持って検討している。例えば、生産性向上や重労働からの解放、科学的知見や治療法等の発見の加速という良い側面に加えて、雇用喪失や人類への破滅的脅威の可能性などについて、学会会議、民間自主組織、政府機関、国際機関など様々な立場から検討や提言が行われている。

「人工知能と人間社会に関する懇談会」では、特に、現存する人工知能技術、または近い将来実現する可能性が高い人工知能技術やそれが普及した社会に焦点を絞り、どのような便益が期待できるか、考慮すべき点は何か、今後取り組むべき課題や方向性は何かを明らかにすることを目的とした。また、人工知能技術に関連して厳密に分離できないデジタルイゼーションの事例についても対象とした。人工知能技術が我々の社会の様々な領域に普及しつつあることを考慮し、移動、製造、個人向けサービス、対話・交流という4つの代表的分野について様々な事例（ケース）を挙げて検討する手法をとった。そして、人工知能技術を使う人、研究・開発する人、文化芸術を担う人、未来の社会をつくる子どもなど様々な世代の人、企業、そして政府など様々な関係者の立場（マルチ・ステークホルダー）から人工知能技術と人間社会に関係する論点を明らかにすることを目指した。

人工知能技術は人間の知的能力と行為を補助し、一部を代替し拡張することを可能とすることから、持続可能社会の強力な推進力になることが期待できる。ただし、以下のような論点について検討し考慮する必要がある。

- ◆ 倫理的論点： 人工知能技術に基づく判断と人の判断のバランスを取ることが大切であり、そのバランスと関係は今後変化していくことが予想され、それに伴い倫理観も変化していくだろう。人工知能技術がもたらすサービスによって、利用者が知らぬ間に感情や信条、行動が操作されたり、順位づけ・選別されたりすることが生じうる場合には倫理的検討が必要である。人工知能技術によって人の認知や行動が拡張されることで人間観がどう変わるかを見ていく必要がある。人工知能技術と協働した創作の価値の受容やその変化、人によって異なるビジョン（人工知能技術との関わり方）や価値観について認識する必要がある。

- ◆ 法的論点： 人工知能技術を使うリスク、使わないリスクを考慮した利活用を推進するために、その利用に伴う法的課題の抽出、例えば事故等によって生じる責任分配の検討や保険の整備が必要である。ビッグデータを活用した人工知能技術の利便性確保と個人情報・プライバシー保護の両立について法的整備・対応が必要である。アルゴリズム開発者やデータ提供者、利用者などの多様なステークホルダーが関与する人工知能技術を利用した創作物等の権利についての検討が必要である。責任概念などの現行法上の基本的な概念について、人工知能技術と社会の相互発展に即した検討と基礎研究が求められるだろう。
- ◆ 経済的論点： 個人にとっては創造的労働が増えるなど業務内容の変化や働き方の変化が生じる可能性があり、それに対応できる能力を身に着けることが望ましい。企業としてはそのような雇用や働き方の変化に迅速に対応することが必要である。政府としては、労働者が業務を変えることを可能とする教育や環境の整備、課題先進国である日本の持続発展可能性のためにも経済格差をなくし、人工知能技術の利活用を促進するための政策が必要である。
- ◆ 教育的論点： 現状の人工知能技術の限界を把握し、協働して創造的活動ができるための能力を身に着けることが望ましい。人の能力を人工知能技術と最大限に差異化し、人にしかできない能力を伸張する教育カリキュラム、従来通り行うべき教育内容の検討などの政策が必要である。
- ◆ 社会的論点： 人工知能技術との関わりの自由について対話する場を作り、多様性や共通性について検討することが大切である。人工知能技術デバインドや人工知能技術に関連する社会的コストの不均衡、人工知能技術によるプロファイリングなど推定結果による差別が生じないような対処が必要である。人工知能技術に対する依存や過信・過剰な拒絶など新たな社会問題や社会的病理が生じる可能性を検討すべきである。
- ◆ 研究開発的論点： 科学者や研究開発者には、倫理観を持ちガイドラインや倫理規定に沿った行動が求められる。セキュリティ確保、プライバシー保護技術などの開発が必要であり、人工知能技術が制御不可能とならないような配慮やその計算過程や論理を説明できる透明性に関する研究が必要である。また、機械学習による確率的な動作に対する社会的受容性を見ていくこと、人工知能技術の多様性の推進、オープンサイエンスの促進、未来社会を設計する人文社会科学研究や融合研究の推進などを通じて、社会に貢献する人工知能技術の研究開発を促すことが重要である。

我々は、検討方法や導かれた論点について国際的な連携や発信を行ってきたが、今後も国際的に協調した取り組みを続けて行く必要がある。特に、人工知能技術と社会の関係検討について国際的な枠組みを考えることが世界全体としての持続可能社会の実現に貢献すると思われるため、世界で共有すべき価値観は何かを検討すべきである。

この懇談会で得られた論点について、社会を構成する全ての人々が自らのこととして受け止め、私たちの未来社会をより良いものとするために具体的な議論を続け、適切な行動を取ることを期待する。

第1章 はじめに

1.1 人工知能技術が社会にもたらすインパクト

1956年、アメリカのダートマス大学で行われた通称ダートマス会議において、「思考する機械」に関する様々な研究を「人工知能」という新しい研究領域とするという提案がなされた。それまでも電子計算機の発明や人の知性を人工的に実現可能かについての研究は行われてきたが、この時点から「人工知能」という言葉を用いた研究領域として設定された。「人工知能」は研究者がその実現を目指している対象であり、未だ完全なものは存在しないし、何をもって完全かということも難しい。そうした研究の過程で、人が知性を用いて行っていると思われている知的活動（認知、推論、学習、思考、これらに基づく行為など）の一部を代替しうる技術が開発されている。それをこの報告書では「人工知能技術」と呼び、主たる検討対象とする。

2016年3月、人工知能技術に基づくプログラムが囲碁の世界チャンピオンに4勝1敗で勝利し、大きな驚きをもたらした。その理由の1つは囲碁が非常に難しいゲームであり、打つ手の可能な範囲が膨大であることから計算機による網羅的な解探索では、世界チャンピオンのような達人には敵わないと思われていたからである。つまり、人にしかできない高度な思考が必要と考えられてきたゲームでさえ、人工知能技術が人を超えたとも言える。また、2014年頃からレベル2と言われる自動運転（補助）機能を備えた車が市販されている。それはカメラやレーダー等の情報を使って加速、操舵、制動のうち複数の機能を自動で行い、高速道路など限定された場面では人と比べても遜色のない運転を実現している。2016年夏には、適切な治療法が見つけれなかったがん患者に対して、人工知能技術を活用することで別の治療法が提案されて回復したという報道がなされた。

また、自動車を資産として持たない企業が情報のやり取りだけでタクシーの機能を実現するライドシェアのようなデジタルイゼーションが、企業活動と我々の働き方や生活の利便性を変えつつある。デジタルイゼーションにおいて用いられる情報の最適化の中には人工知能技術と厳密に分離できない事例も少なくない。

現在、日本政府は少子高齢化などの社会課題を科学技術によって克服し、未来に向かって持続発展可能なSociety 5.0¹の実現を目指している。そこで、人工知能技術はロボットやデジタルイゼーションを加速するIoTなどと共に重要な基盤技術とされている。

このように近年の人工知能技術の進展は社会に大きな驚きをもたらしており、政府や企業は持続可能社会実現への原動力として、また産業力強化の中核技術として期待し、研究開発に多大な投資をしている。

¹ Society 5.0 とは、「必要なもの・サービスを、必要な人に、必要な時に、必要なだけ提供し、社会の様々なニーズにきめ細かに対応でき、あらゆる人が質の高いサービスを受けられ、年齢、性別、地域、言語といった様々な違いを乗り越え、生き活きと快適に暮らすことのできる社会」であり、人々に豊かさをもたらすことが期待される（第5期科学技術基本計画）

<http://www8.cao.go.jp/cstp/kihonkeikaku/index5.html>（2017年2月24日取得）。

1.2 人工知能技術がもたらす期待と不安

人工知能技術は人間社会に多大な便益をもたらすと期待される。例えば、自動運転車は人の不注意やミスで起きている事故を減らし、安全性を高めると言われている。そして、様々な機械の自動化が実現すれば人は煩雑な作業から解放され、時間的余裕ができ、より価値の高い仕事や趣味に時間を割くことが可能となる。がんの治療法提案のように、専門家の知識や経験が必要で、かつ膨大な時間がかかる知的作業が迅速化され、高効率で知的発見が可能となる。したがって、人工知能技術は、私たちの社会をより便利に、豊かに、安全安心にすることに貢献することが期待される。

一方で、人工知能技術の発展に不安を感じている人も存在する。自動運転車が事故を起こした場合誰が責任を取るのか、人工知能に仕事を奪われてしまうのではないか、人工知能技術が普及したら子どもたちは学習する意欲を失うのではないか、あるいは、もしかしたら人は人工知能に操られてしまうのではないか等の疑念・不安が生じうる。この報告書の目的は、これらの疑念・不安に対して性急に安全性や安心を保証することではなく、現状を整理し、何が問題か、何を今後検討していくべきかを明らかにすることである。現時点でほとんど考慮する必要のない心配もあるが、一方でいくつかの不安要素に関しては現状あるいは今後の対応方法を検討していくことが個人（人工知能技術を利用する人も研究開発する人も）としても、企業としても、政府としても必要であろう。

社会への人工知能技術の導入や利活用には、短期的には一定の懸念や対処が必要な可能性もあるが、懸念やリスクに適切に対処することで人工知能技術から得る便益はより大きくなる。成熟した社会である我が国は課題先進国とも言われている。経済成長、少子高齢化や地域社会の自律的発展、安心安全で豊かな生活の実現、自然災害対策などの課題を解決し、持続発展可能な社会を推進するためにも人工知能技術の貢献が期待される。

1.3 人工知能技術はこれまでの技術と何が違うのか

これまでも新しい技術の登場に対して我々は期待し、時には不安を感じ、それらに対応して技術を使いこなして社会に受容してきた。今なぜ特別に人工知能技術に関して人間社会に及ぼす影響を考慮する必要があるのだろうか。

人工知能技術がこれまでの他の技術と最も違う点は、「（人のみが持っている）知性を用いないとできないと思われていたこと」を部分的ではあるが代替する技術だということである。そして、人工知能技術は適切な目標を与えればデータから自ら学習することができ、徐々に改良していくことで急速に高度化する。さらに、人工知能技術は、電気機器や自動車、携帯電話のように目に見える外形のある技術ではなく、すでにある様々なモノやサービスの内側で機能する技術であるという点も重要である。プログラムの一部として機能することから、そこに人工知能技術が利用されているかどうか、どの程度利用されているのかが利用者からは分かりにくいことが多い。これまでは人が道具や技術を使っていたが、人工知能技術も人のように既存の道具や技術をほぼそのまま利用することが可能である。つまり、外見からは人工知能技術利用の有無は区別がつかない。したがって、社会の様々な技術やシーンの中に既に人工知能技術が使われている可能性があり、あるいは人工知能技術の利用を謳っていても実際には限定的にしか使われていないか全く使われていない可能性すらある。つまり、人工知能技術は、その開発速度が速く外からはその動作が見えにくい技術であることが多いことから、知らない間に我々の周りに普及し高度化していくという可能性も考えう

る。

ゆえに、従来型の技術革新に対して行ってきた対応に加えて、上記で述べたような人工知能技術の特徴に応じた対応が必要であろう。

1.4 懇談会の目的

「人工知能と人間社会に関する懇談会」²は、内閣府特命大臣（科学技術政策担当）の下に、日本が目指す社会Society 5.0を実現する推進力となる人工知能技術の研究開発と利活用を健全に進めるために2016年5月に設置された。特に、現存する人工知能技術、または近い将来実現する可能性が高い人工知能技術やそれが普及した社会に焦点をあて、どのような便益が期待できるか、考慮すべき点は何か、今後取り組むべき課題や方向性は何かを明らかにすることを目的とし活動を行ってきた。

人工知能技術が我々の社会の様々な領域に普及しつつあることを考慮し、移動、製造、個人向けサービス、対話・交流という代表的な4つの分野について様々な事例（ケース）を挙げて検討する手法をとった。また、人工知能技術に関連して厳密に分離できないデジタルイゼーションの事例についても対象とした。そして、人工知能技術を使う人、開発する人、未来の社会をつくる子どもを含む様々な世代の人々、文化や芸術を担う人、企業、そして政府など様々な関係者の立場（マルチ・ステークホルダー）から、人工知能技術と人間社会に関係する重要な論点を明らかにし、今後継続的に行う検討の端緒となることを目指した。

² <http://www8.cao.go.jp/cstp/tyousakai/ai/index.html>（2017年2月24日取得）

第2章 人工知能と人間社会に関する検討動向

人工知能技術や知的機械 (intelligent machine) の急速な進展に伴い、その人間社会への影響について検討する動きが世界的に同時にいくつも起こっている。特徴的なことは、人工知能研究に関係する科学者、工学者そして政府機関のみならず、他分野の研究者や企業家、そして一般市民も強い関心を持っていることである。

2.1 世界の動向

2008-2009年にアメリカ人工知能学会 (Association for the Advancement of Artificial Intelligence; AAAI) では学会長が主導する有識者会合を作り、人工知能の発展が社会に及ぼす長期的影響について検討した (2008-2009 AAAI Presidential Panel on Long-Term AI Future)。2014年にその発展形として、スタンフォード大学 (アメリカ合衆国) に「One hundred year study of Artificial Intelligence」という活動が発足した。人工知能の発展が社会に及ぼす長期的影響について好機と懸念 (opportunities and concerns)、法律や倫理の問題などを研究者中心の構成員が継続的に調査・検討している。2016年のレポート³では、2030年までの典型的な北米都市を想定して交通、家庭、ヘルスケアなど分野ごとに人工知能研究の影響について検討を行い、政策提言も行っている。また、2030年までには人類が人工知能に滅ぼされるような懸念はないと明言している。標準規格の制定も行う米国電気電子学会 (IEEE) は、2016年12月に100名以上の人工知能や倫理の有識者の意見を集めて「Ethically Aligned Design, Version 1」を公開した⁴。その目的は、自律的な知能システムの全ての開発者に倫理的な視点や考慮すべき重要課題を提案すること、および倫理指針の標準化への提言である。

一方、ボストン (アメリカ合衆国) に2014年に発足した民間自主組織「Future of Life Institute (FLI)」は、人文社会科学や自然科学など様々な分野の研究者のみならず、俳優、学生、企業家など多様な背景をもつ人が集まり、人工知能の発展を中心にバイオテクノロジー、核兵器、気候変動が人類にもたらしうる危険性 (Existential Risks) を扱う。寄付金を原資とした社会に良い人工知能技術のための研究助成や研究会⁵の開催、啓蒙活動等を行っている。汎用人工知能 (Artificial General Intelligence; AGI) の将来的危険性や人工知能に関する安全性 (AI Safety) の問題も扱う。

オックスフォード大学 (英国) の「Future of Humanity Institute (FHI)」 (2005年創設) では、人工知能の安全性も含めて人類の存亡に関わる危険性 (Existential Risks) が研究されている。ケンブリッジ大学の「Centre for the Study of Existential Risks (CSER)」 (2012年創設)

³ Peter Stone, Rodney Brooks, Erik Brynjolfsson, Ryan Calo, Oren Etzioni, Greg Hager, Julia Hirschberg, Shivaram Kalyanakrishnan, Ece Kamar, Sarit Kraus, Kevin Leyton-Brown, David Parkes, William Press, AnnaLee Saxenian, Julie Shah, Milind Tambe, and Astro Teller. "Artificial Intelligence and Life in 2030." One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel, Stanford University, Stanford, CA, September 2016. Doc: <http://ai100.stanford.edu/2016-report> (2017年2月24日取得)

⁴ http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html (2017年2月24日取得)

⁵ Beneficial AI 2017. <https://futureoflife.org/bai-2017/> (2017年2月24日取得)

では、汎用人工知能を中心として科学技術の発展によって生じうる世界的な破滅的危機（Global Catastrophic Risks）の予測や影響評価がなされている。

企業においても人工知能研究に関する倫理委員会（Institutional Review Board: IRB）を設置する動きや、複数企業がパートナーシップを結び人工知能が社会や倫理に悪影響を及ぼさない方法を共有し、社会に及ぼす影響について正しく広報する仕組みを作ることが報道されている。

政府機関や国際機関でも同様の検討が始まっている。例えば、アメリカ合衆国ではホワイトハウス科学技術政策局（White House Office of Science and Technology Policy; OSTP）が一般市民も参加するワークショップを2016年5月から4回実施し、人工知能技術が社会にもたらす便益とリスクについて検討と対話を行った（同年10月に報告書公開⁶）。経済協力開発機構（The Organisation for Economic Co-operation and Development; OECD）ではデジタル経済政策委員会（The Committee on Digital Economy Policy）が中心となり、人工知能技術も含むデジタルライゼーションが社会にもたらす便益とそのために必要な政策を検討する試みを始めた。

2.2 日本の動向

日本の人工知能学会は、学会誌の名称と表紙の変更が社会的に大きな話題になったことも契機となり、一般社会における人工知能の高いプレゼンスと強いインパクトを考慮して、社会との接点を重要視するようになった⁷。そして、人工知能と社会との関わりについて議論、考察し社会に発信することを目的に2014年12月に倫理委員会が発足し、2017年2月には倫理指針が公開された。科学技術社会論や人工知能の研究者が集まる「AIR (Acceptable Intelligence with Responsibility)」研究会は人工知能の社会的受容性に関して現場視察やインタビューを行うなど対話を中心として検討している。情報ネットワーク法学会のロボット法研究会は社会制度の国際的研究動向と消費者保護について検討を行っている。このように様々な視点からの取組が行われている。

総合科学技術・イノベーション会議が策定し2016年から始まった「第5期科学技術基本計画」⁸は、長期的視野に立ち体系的かつ一貫した科学技術政策を実行するためのものである。目指すべき社会としてのSociety 5.0の実現のために人工知能技術も重要な役割を担うことに加え、科学技術イノベーションと社会との関係深化の重要性、そのために倫理的・法制度的・社会的取組を行うべきと示されている。また、総務省は2016年2月から「AIネットワーク化検討会議」を設置し、人工知能のネットワーク化が社会・経済にもたらす影響やリスクの評価、今後の課題の検討を行い、研究開発の原則を公表した⁹。2016年10月からは「AIネットワーク社会推進会議」として引き続き検討の具体化を進めている。

6

https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf（2017年2月24日取得）

7 松尾ら(2015)、人工知能学会倫理委員会の取組み、人工知能、30(3)、358-364。

8 第5期科学技術基本計画 <http://www8.cao.go.jp/cstp/kihonkeikaku/index5.html>（2017年2月24日取得）

9 「「AI ネットワーク化検討会議」中間報告書の公表」http://www.soumu.go.jp/menu_news/s-news/01iicp01_02000049.html 2017年2月24日取得。「AI ネットワーク化検討会議 報告書2016の公表」http://www.soumu.go.jp/menu_news/s-news/01iicp01_02000050.html（2017年2月24日取得）。

人工知能は人間を超えるか？

サイエンスフィクション（SF）の世界では人間と同等またはそれ以上の知能を持ち、人類の脅威として、または人間と同じように葛藤する存在として人工知能等が描かれてきた。昨今、人工知能と人間社会に関して世界中で議論されているが、その一部は、そうしたSFにおける人工知能像の流れを汲んでいるものもあるのではないかと。以下では、人工知能と人間社会に関して今後さらに議論を深める上でのヒントとすべく、SFにおける人工知能の例を取り上げ、その蓋然性等について考察を試みる。

人類の管理や滅亡を企む人工知能は映画「ターミネーター」や「マトリックス」、漫画「火の鳥 未来編」等多くのSFで描かれてきた。映画「2001年 宇宙の旅」や「エイリアン」、「ブレードランナー」、「エクス・マキナ」等では人間に危害を加えようとする人工知能が登場した。

人間と同じように葛藤する人工知能も映画「アンドリューNDR114」や「A. I.」等で描かれてきた。映画「Her」では人工知能と人間の男性との恋愛が描かれている。こうした人工知能については、機械とプログラムから構成される人工物でありながら人間と同じく心を持つように見える異質な存在とどう関わっていくべきか、そもそもそのような人工知能を開発してもよいかといった議論がある[1]。

このような高度な知能を持つ人工知能の出現の可能性については、人間のクイズ王[2]や碁のチャンピオン[3]、ベテラン戦闘機パイロット[4]に勝利するだけでなく、料理[5]や小説[6]といった創作に関する領域でも人間に迫ろうとする等、最近の人工知能研究の急速な進展を踏まえると一定の説得力があるように思える。

しかし、現実の人工知能技術は全てコンピュータプログラム、即ち、設計者の目的を具体的な計算手順として表現したものである。深層学習をはじめとする近年の進歩により高度な学習機能を備えるようにはなったが、上述したSFにおける人工知能と違い、自ら目的を設定することはできず、基本的には設計者の意図通りに動作するのみである。機械学習や統計的手法に基づく人工知能技術は時に設計者の意図に沿わない動作をすることがあるし、バグがあれば誤動作はするが、設計者が意図しなかった目的を自ら設定し、その達成に向けて合理的な動作をする可能性はゼロに近い。現実の人工知能技術に与えられる目的とは、所与の入力から設計者が望む出力を得るといった特定の具体的な課題に他ならず、1つの人工知能技術が為しうるのは画像や音声の認識や自動走行といった個別具体的な課題に留まる。現存する人工知能技術は人間と同じような汎用的な知能を持つのではなく、個別具体的な課題を自動的に遂行するに過ぎない。

今後の数多くの劇的なブレイクスルーにより、現存する技術の延長線上にはない、従来と全く異なる原理に基づくハードウェアやソフトウェアが登場することで、そうした限界を打ち破り、汎用人工知能あるいは Superintelligence（あらゆる点で最も頭脳明晰な人間を上回る知能を持つ仮説的人工知能 [7]）が出現し、人類への脅威となる危険性は完全には否定できない。また、人工知能技術を悪用する人間が社会に害を及ぼす危険性については、他のあらゆる技術と同様、現時点でも十分注意を払う必要がある。しかし、人工知能が自ら目的を持ち、人類に危害を加え

るという SF のような世界が現実のものとなる蓋然性は少なくとも今後数十年は極めて低いと言
ってよいのではないか。

2016 年 10 月に公開された米国ホワイトハウスの報告書「Preparing for the Future of Artificial Intelligence」[8]にも「汎用人工知能は少なくとも今後数十年の間は実現しないだ
ろう」と書かれている。One Hundred Year Study on Artificial Intelligence (AI100) [9]では
スタンフォード大学が中心となって人工知能と人間社会に関する議論を深めているが、2016 年 9
月に出された報告書「Artificial Intelligence and Life in 2030」[10]には「大衆紙に書かれ
ているような予測とは異なり、人工知能が人類の脅威となるような要因は見つからなかった」と
書かれている。機械学習を専門とする Baidu 社の Andrew Ng 氏は、人類に危害を加えるロボット
について心配するのは火星での人口過剰を心配するようなものだと述べている [11]。

一方、人工知能を始めとする「人類滅亡の危機 (Existential Risks)」について研究している
Future of Life Institute (FLI) [12]は、中心的な研究対象を人間と同等の知性を持つ人工知能
又は Superintelligence としている理由について、「そうした人工知能が人類に与える影響の甚
大さを考慮すれば、今から研究すべきである」と述べている [13]。

今後、人工知能と人間社会について議論を進めていく上で重要となるのは、上記 2 つの対極的
なポジションのバランス、即ち、技術レベルの進展を正確に把握した上での地に足の着いた現実
的な議論と、蓋然性に関わらず常に最悪の事態も想定した議論のバランスを如何に取るか、では
ないだろうか。

参考文献・URL

1. 松尾 豊, 西田 豊明, 堀 浩一, 武田 英明, 長谷 敏司, 塩野 誠, 服部 宏充, 江間 有沙, 長
倉 克枝. 人工知能と倫理. 人工知能, 31(5), 635-641. (2016)
2. スティーヴン・ベイカー (著), 土屋 政雄 (翻訳). IBM 奇跡の“ワトソン”プロジェクト:
人工知能はクイズ王の夢をみる. 早川書房. (2011)
3. <https://deepmind.com/research/alphago/>
4. http://www.huffingtonpost.jp/engadget-japan/ai-fighter_b_10746878.html
5. http://www.huffingtonpost.jp/2014/07/08/chef-watson_n_5568914.html
6. http://www.fun.ac.jp/~kimagure_ai/
7. Nick Bostrom. Superintelligence: Paths, Dangers, Strategies. Oxford University
Press. (2014)
8. https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf
9. <https://ai100.stanford.edu/>
10. <https://ai100.stanford.edu/2016-report>
11. <https://www.wired.com/brandlab/2015/05/andrew-ng-deep-learning-mandate-humans-not-just-machines/>
12. <http://futureoflife.org/>
13. <http://futureoflife.org/2015/10/12/ai-faq/>

第3章 懇談会におけるアプローチ

本懇談会では、遠い未来の物語や常識的にはありえない不安について抽象的な議論をするのではなく、地に足の着いた具体的な議論を目指した。そのため、「現存する人工知能技術、あるいは近未来に実現することが確実な人工知能技術」を対象とし、それらが社会にもたらす影響を検討するため、「事例に基づくアプローチ」を採用した。まず、いくつかの分野について近未来像を想定し、倫理、法、経済などの観点から、様々な具体的事例を列挙した。その際、有識者である懇談会構成員や各分野の専門家、実際に企業活動を行う経営者などの意見に基づき検討した。また、幅広い意見や事例を集めるために、インターネットのホームページを使った対象を限定しない意見募集や人工知能と人間社会の関係に興味のある人を集めて行ったブレインストーミング的ワークショップ、日本科学未来館での子どもを含む一般来場者を対象としたワークショップも行い参考にした。そして、複数の事例に共通する論点を抽出し整理した。このようにすることで、例えば雇用の問題についても経済的観点のみならず教育的観点や研究開発的観点などからも検討することが可能となり、人工知能技術と社会の関係についての多面的な検討が行われた。

3.1 分野の選択

人工知能技術は、近未来の生活の様々なところに入ってくると予想される。そこで、生活にとって身近であることや日本の産業力の強みなどを考慮して、4つの分野「移動」、「製造」、「個人向けサービス（医療、金融を含む）」および「対話・交流（コミュニケーション）」を検討対象として選択した。これらは、人工知能技術が適用される個々の技術を意味するのではなく、技術を活用する分野を意味し、そこに含まれる様々な事例を検討した。

- 移動： 人や物の移動に関わる分野。事例としては、自動運転車、ライドシェア、ドローン配送など。
- 製造： 物の製造や作品の制作に関わる分野。事例としては、自動化した製造ラインを利用するスマート工場、人の作業能力を補綴・拡張するロボットスーツ、人工知能技術による精緻な製造技術、芸術家や達人の作風や技術の複製など。
- 個人向けサービス（医療、金融を含む）： 個人に関する情報サービス分野。事例としては、インターネットショッピングや検索での推薦サービス、人工知能技術を活用する健康・医療サービス、ロボアドバイザーやアルゴリズム取引、与信審査への人工知能技術応用など。
- 対話・交流（コミュニケーション）： 人同士や人と機械とのコミュニケーションに関わる分野。事例としては、質問応答のための対話エージェント、チャットボット、機械翻訳など。

3.2 検討する観点

科学技術と社会との関係については、ELSI (Ethical, Legal, and Social Issues) という用語があり、ヒトゲノムプロジェクト以来、倫理的、法的、社会的観点について検討することが求められている。本懇談会では、人工知能技術が産業や労働に及ぼす影響の大きさを考慮して経済的観点、社会の変化と未来社会の構成員となる子ども達への影響を考慮して教育的観点、人と社会に関する検討内容を人工知能技術の研究開発に反映するために研究開発的観点を加え、6つの観点から論点を検討することとした。

- 倫理的論点： 人工知能技術のもたらす結果に対する倫理的側面や道徳性、価値などについて。
- 法的論点： 人工知能技術を利活用する際に関連する法律や契約、人工知能技術のもたらす結果に対する責任、個人情報保護などについて。
- 経済的論点： 人工知能技術を利用した経済活動やそれを利活用する際の働き方についての被雇用者個人、企業、そして政府からみた配慮すべき点などについて。
- 教育的論点： 人工知能技術が利活用されるときに教育はどう変わるのか、教育を提供する側や教育を受ける側の視点からの問題などについて。
- 社会的論点： 様々な人と人工知能技術との関わり方、人工知能技術の利活用によって副次的に生じうる社会問題などについて。
- 研究開発的論点： 人工知能の研究開発者が注意すべきことや、利用者が人工知能技術の研究開発にどう対処すべきかなどについて。

必ずしもこれらの観点到厳密に分けられない問題もあり、倫理と法律など複数の観点到跨がる論点もあるが、便宜的にこれらの観点到整理を行い、複数の観点到跨るものについてはその旨言及した。

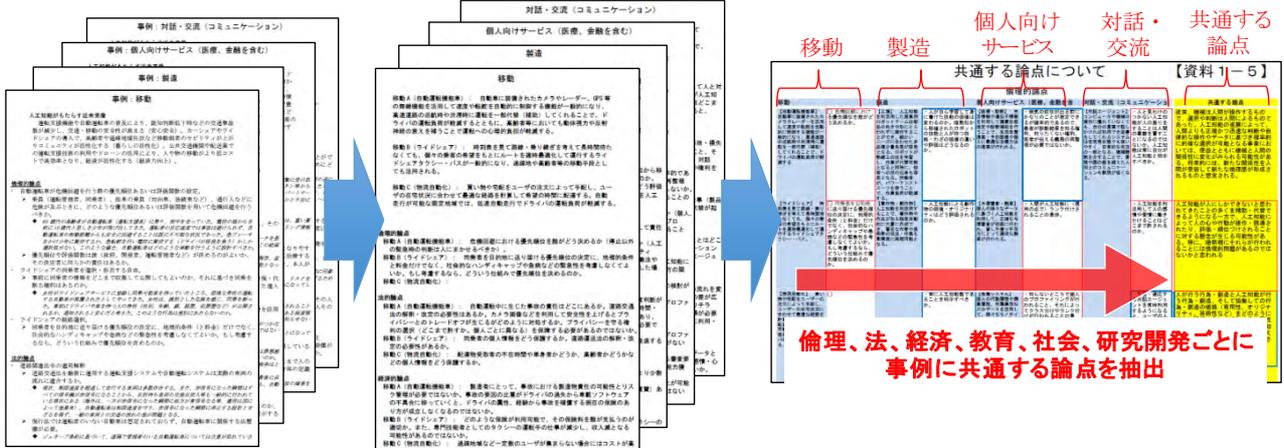
3.3 共通する論点的整理へ向けて

上記のように4つの分野「移動」、「製造」、「個人向けサービス（医療、金融を含む）」および「対話・交流（コミュニケーション）」について、多くの事例を具体的に検討し、倫理的、法的、経済的、教育的、社会的、研究開発的論点についてそれぞれ検討した後に、共通する論点的整理を目指した。そのために、各分野での近未来の社会実装例を代表的な1-3つに絞り込み、それらに関する各事例と分野とからなる表を作成した。次に、観点到ごとに（例えば倫理的論点について）、横串を差すように個別事例を見ていき、分野をまたがって共通している論点进行考察し整理した。

分野別の事例の深掘り検討

論点の絞り込み

共通する論点の抽出



論点整理のための分野別事例の表（右）の導き方

この方法を用いることで、抽象的な論点のみを議論するのではなく、地に足の着いた、現実的な問題あるいは近未来に具体的に想定できる問題として論点を抽出することが可能となった。ただし、世の中の事例をすべて列挙することは不可能であり、こうして抽出した論点が必ずしも全てを網羅しているわけではないことに注意する必要がある。事例によってその時間軸や規模には多少の違いがあり、例えば悪意を持つ者による人工知能技術の利用、人工知能技術の想定外・目的外の利用、人工知能技術が自律性を持ち相互に協調して社会に影響するような事態などに対する検討は十分とは言えない。こうした問題については、分野専門的な組織において今後の検討が望まれる。本懇談会では実現性や重要性の高い事例に焦点を絞り、検討の端緒として論点をとりまとめた。

第4章 人工知能技術と人間社会について検討すべき論点

人工知能技術は人間の知的能力と行為を補助し、一部を代替し、拡張することを可能とする。したがって、人工知能技術は私たちの社会に多大な便益をもたらす、健全に利活用されれば持続可能社会の強力な推進力になることが期待できる。ただし、その実現のためには人工知能技術と人間社会の適切な関係について前もって考え、準備をしておく必要がある。そのために今現在検討し考慮すべき論点をまとめた。

4.1 倫理的論点

人はこれまでも様々な道具や機械を利用して、状況に応じた選択や判断を行い、行動してきた。現在、人工知能研究の進展によって、膨大なデータに基づいた迅速で正確な判断、半自動的な操作、統計に基づく的確な選択が可能となることが増えてきた。人工知能技術が人の選択や判断を支援することで、その正確さや迅速さは向上し、人が犯しやすい認知バイアスや偏見¹⁰の影響を受けない判断が可能になるなど多くの便益がもたらされる。ただし、必ずしも全てを人工知能技術に委ねるのではなく、判断する状況や対象に応じて、人による判断と人工知能技術に基づく判断のバランスを考慮する必要がある¹¹。人工知能技術がさらに進展していくと、人工知能技術・機械と人間との関係性に徐々に変化がみられる可能性がある。将来的にはその新たな関係性に基づいて、新たな倫理観が形成されることも予想される。**(人工知能技術の進展に伴って生じる、人と人工知能技術・機械の関係性の変化と倫理観の変化)**

人工知能技術は、人にしかできないと思われてきた高度な思考や推論、行動を補助・代替できるようになりつつある。その一方で、人工知能技術を応用したサービス等によって人の心や行動が操作・誘導されたり、評価・順位づけされたり、感情、愛情、信条に働きかけられることがあるとすれば¹²、そこには不安や懸念が生じる可能性がある。特に、本人が気づかないところでそれらが行われる場合には、倫理的検討が必要であろう。**(人工知能技術によって知らぬ間に感情や信条、行動が操作されたり、順位づけ・選別されたりする可能性への懸念)**

¹⁰ 確証バイアス（自分の仮説に都合のよい情報を過剰評価し、確証に重きを置く認知的傾向）や根本的帰属エラー（人の行動の結果を状況・環境よりも個人特性のせいにする認知的傾向）などの認知バイアス、および個人的背景や文化的背景に基づく偏見など。ただし、機械学習に基づく人工知能技術は学習データの偏りによって出力される判断が偏ることも知られており（アルゴリズムバイアス）、そのバイアスによる不公平が生じることを防ぐための研究が求められる。

¹¹ 例えば緊急事態に運転者1名の命と対向車の多数の乗員の命のどちらを救うかを功利・効用に基づいて判断可能かは現状では解けない倫理的問題である。人は自分の命を守ろうとするかもしれないし、人工知能技術では効用に基づいて多数の命を守るように設定することも可能である。そこで現段階では、人と人工知能技術の判断のバランスをどうするかが問題となる。

¹² 古くから広告等で同様の懸念があり、一定の自主規制や注意喚起がなされている。人工知能技術を用いる場合には、サービスや広告が個人特性に応じて最適化されることからより細やかな注意が必要であろう。

人工知能技術は、人のこれまでの時空間感覚や身体感覚を拡張¹³する。それに伴い、人の能力についての概念や感情の捉え方についても変化が生じる。これらの相互作用もあり、その受容のために人間観の捉え直し¹⁴が行われていく可能性もある。**(能力や感情を含む人間観の捉え直し)**

人工知能技術の利活用によって、生産性が向上する。たとえば、これまで芸術家や達人しか創れなかったものや膨大な時間とコストが必要だったものが、容易・安価に生産され、誰もが利用できるようになる。そのとき、人が主として行った行為・創造、人工知能技術が主として行った行為・創造、そして人と人工知能技術が協働して行った行為・創造の価値（有用性、オリジナリティ、芸術性など）がそれぞれどのように評価されるのか、社会的に受容されていくのかについて、注意して見ていく必要がある。また様々な人が対話して考える場を作ることも大切である。特に、人と人工知能技術が協働することは人間能力の拡張とも言え、新しい価値観の基盤となる可能性がある。ただし、人によって人工知能技術や機械に関する価値観や捉え方は違うことを認識し、様々な選択肢や価値の多様性について検討することが大切である。**(人工知能技術が関与する行為・創造に対する価値・評価の受容性。価値観や捉え方の多様性)**

4.2 法的論点

人工知能技術が積極的に利活用され社会に安心して受容されるためには適切な法的検討が必要である。例えば、交通事故の大多数は人のミスや不注意によって生じると言われており、自動運転車が増えることで統計的には事故が減少し、より安全な社会になると期待される。しかし、自動運転車が事故を起こしたときにその責任はどこにあるのかという問いが生じる。このような社会実装が近い具体性の高い人工知能技術については、それがもたらす便益や成果および人工知能技術がもたらすリスクや事故、権利侵害等について、その責任の分配を明確にすることが必要である。特に、技術進展のレベル（たとえば、自動運転のレベル0～4のそれぞれなど）に対応した責任分配を明確にし¹⁵、それ以外の不確実で確率的に生じるようなリスクに対しては保険¹⁶を整備して対応することが、人工知能技術が社会に受容され、その便益が享受されるために有効であろう。産業界における萎縮効果やレピュテーションリスクへの過剰反応を防ぐためにも責任分配の明確化と保険の整備は重要である。そして、人工知能技術を利用することによるリスクのみならず、利用しないことで便益を失うリスクや責任¹⁷も意識して人工知能技術の利活用を検討するべきであろう。**(人工知能技術による事故等の責任分配の明確化と保険の整備。人工知能技術を使うリスク、使わないリスクの考慮)**

人工知能技術はビッグデータの活用でより有益となる。その利便性と個人情報保護（プライバシー）を両立し、萎縮効果を生まないための制度（法律、契約、ガイドライン）の検討が必要であ

¹³ 自動運転車など自動化した機械や身体補綴技術の日常的利用に基づく変化など。

¹⁴ 人の能力とは何か、身につけている人工知能技術や補綴技術を含めてその人の能力なのか、など。

¹⁵ 自動運転のレベル2までについては、交通事故の責任は基本的に運転者にあるとされている。

¹⁶ ただし、政府が社会保険のような強制保険を設けるべきか、保険料を誰が負担するかなどの検討が必要である。

¹⁷ たとえば人工知能技術を活用した診断で治療法を変えれば回復したかもしれない患者が、そのような診断を受けられなかった場合など。

る。個人情報についてのデータアクセス権、データポータビリティ、そしてそのセキュリティなどについて、国際的に協調して議論していくことが重要となることから、日本でも今後検討していくことが必要である。また、それを実際に適用する例として行政サービスへの人工知能技術活用を政府が検討することが期待される。**(個人情報とプライバシーの保護も含めたビッグデータ利活用)**

人工知能技術の活用によって高付加価値な創作物が容易に生産されるようになると予測される¹⁸。人工知能技術を利活用した製造では、アルゴリズム開発者、学習データ提供者、サービス提供者、実際に製造した人など多くの人が関わる。そこで、人工知能技術による創作物や人工知能技術と人が協働した創作物、学習モデルなどの権利は誰のものかなどのいわゆるデータオーナーシップの検討が必要である。その際、人工知能技術の開発と利活用を促進するために、排他的な権利保護ではなく、状況に応じた契約とガイドラインによって、アルゴリズムの開発者や活用者、学習データ提供者等へ適切な権利（インセンティブなど）を配分する方法を検討することも有効であろう。

(人工知能技術を活用した創作物等の権利の検討)

人工知能技術の進展による社会や雇用の変化に対して、従来法（道路関連法、医薬品医療機器法、製造物責任法、労働関連法¹⁹など）の解釈で対応が十分か、法律の改正が必要か、新しい法律が必要か、あるいは法律の概念自体を検討する必要があるかを検討し、議論を続けていく必要がある。例えば、機械学習による人工知能技術が社会実装されるにつれ、その判断や人と人工知能技術が協働した行為によって生じる結果についてその責任の所在を明確にするのが難しい事象が生じつつある²⁰。人工知能技術が社会に受容されていく過程において、技術のみならず人間社会も相互同時に発展していくことが必要であり、責任とは何かという問題などに関する現行法上の基本的概念についての検討や基礎研究が望まれる²¹。**(法解釈、法改正、法に関連する基本的概念の再検討の可能性)**

4.3 経済的論点

人工知能技術により経済・産業が活性化するとともに、機械学習に必要なデータを生成するためなどに新しい雇用が生まれる可能性がある。いわゆる人工知能覇者企業が出現して、既存のビジネス勢力図が抜本的に変化する可能性があり、このような産業的独占について社会との関係への影響も考慮して見ていく必要がある。多くの企業にとって、人工知能技術を活用することで多大な労働力を必要とせずとも大規模な経済活動が可能となり、コストメリットが高く機動力も上がることが

¹⁸ 熟練工と同等かそれ以上の作業を行う産業用ロボットや著名な芸術家の作風を学び創作する人工知能技術などが実用化されつつある。

¹⁹ 例えば、経済的論点で後述するように個人事業主化が進むと、被雇用者を対象とする労働法について捉え直しが必要となる可能性がある。

²⁰ 例えば、筋電位の変化を検出して人の身体動作の意思を推測して動くロボットスーツが他人を傷つける事故を起こした場合、その筋電位は人の意思を反映しており使用者に責任があると考えられるのか、そのような推測結果をもたらした学習データ提供者あるいはアルゴリズム開発者に責任があるのかは切り分けが非常に難しい問題である。

²¹ 例えば、自動走行車については統計的に交通事故数が減り安全性が上がるが、その中でまれだが学習データから乖離した状況などにおいて責任が明確ではない事故が生じる可能性がある。そのような事故についての経済的損失は保険で対処することが可能であろう。しかし、このような事象を社会としてどう受け入れるかは、法学や倫理学、社会学等の基礎的・本質的問題であり、幅広い議論が必要である。

予想される。一方で、人工知能技術が倫理・法・社会などとの整合性を作り上げていく過渡期には、逆に経済的に非効率な状況が生じる可能性もあることから、迅速で適切な対応が必要である。

被雇用者にとっては、現在の仕事・業務内容（タスク）によってはその役割が人間から人工知能技術・機械に代わり、人は単純労働から解放され、より創造的な業務を主に行うようになることが予想される。ただし、今の業務から新しく必要とされる業務に移行することが難しい場合、つまり人材と業務のマッチングがうまく行われない場合には、失業と人手不足の両方が同時に生じる可能性がある。そのため、労働者は個人の能力を人工知能技術と差異化して伸張させること、その能力を最大限に発揮するために仕事を変える転職力や創造的労働に必要な能力、人工知能技術を活用する能力などの獲得を主体的に行うことが望まれる。今後、人工知能技術を活用した起業なども増加し、個人事業主化が進むことが予想される。（人工知能技術による業務や働き方の変化：個人対象）

企業の国際競争において人工知能技術は欠かせないものとなっている。人工知能技術の利活用によって単純労働や長時間労働が減少し、高付加価値な労働と企業に従属しない自由な働き方への変化が予想される。それに合わせてかつ人工知能技術の進展速度の速さも考慮して、企業は、経営判断の迅速化、雇用の再配置の迅速化、テレワークなど空間と時間に制約されない働き方の促進を検討して持続的な経済成長を目指すことが必要である。特に人工知能技術を開発する人材や使いこなせる人材の確保、あるいは他の業務からの配置転換が必要となった時に迅速に対処できる機動力が必要である。配置転換のためには業務と並行した職業訓練や学習機会の提供が有効であろう。（人工知能技術の利活用による雇用と企業の変化：企業対象）

国の政策としては、人工知能技術を活用して経済成長を促し、個人に適した多様な労働形態を確保することが必要である。なにより労働移動²²を可能とする能力を育成し、学習する機会を提供すべきである²³。求められる業務と人材とのマッチングを促進する仕組みも求められる。これら教育政策と雇用政策は連動させることで効果が最大化する。それに加えて、何らかのマクロ経済政策やセーフティネットが必要かを検討することが大切であろう。人工知能技術による生産効率向上や経済活性化、予測可能性の向上などの恩恵・利益をどのように社会的に公平に配分し、経済格差をなくすかを検討する必要がある。そして、労働力不足に直面する日本にとっては特に人工知能技術の有効性が高いことから、産業競争力を向上させる政策をさらに加速することも有効であろう。もちろん、利用者側も自ら考えて産業や政策に期待することや考慮すべき点を要求することが大切である。（人工知能技術の利活用を促進するための経済政策、労働移動を可能とする教育政策・雇用政策：国対象）

4.4 教育的論点

人工知能技術の利用者はその便益とリスクを理解し、責任の所在を見極め、人工知能技術によってどのような選択や操作がなされているかを理解して使いこなすことが望ましい。つまり、人工知能技術の優位点と限界を把握し、協働・協調して創造的活動ができるような能力を身に着けること

²² 労働者個人が、労働市場において企業間、産業間、職業間などを移動すること。

²³ これは教育的論点にも共通する問題であり、子どもへの教育に加えて、働きながら常に知識とスキルを更新するための生涯教育の重要性が増す。

が望まれる。(人工知能技術を適切に利活用するための教育)

教育政策において求められるのは、現状の人工知能技術では何ができないかを調べ、そのエビデンスに基づいた教育カリキュラムを検討することである。例えば、データの統計的処理に基づくのみでは未だ難しいとされる深い意味理解、自らの実体験に基づいて想像力を働かせ未知の世界をより深くイメージできる力、解決すべき問題そのものを見つけ出す能力、共同作業のためのコミュニケーション能力、能動的に新しい情報を探求し異なる意見を取り込み検討する能力などの重要性が増すと考えられる。それによって、人の能力を人工知能技術から最大限に差異化することが可能となり、人が人工知能技術を利活用して創造的業務を行い、少ない労働力でも高い生産性を実現する社会を目指せる。なお、人工知能技術の進展の速さに対して子どもの教育には時間がかかることから、特に学校教育の検討は急務である。また、人工知能技術で代替可能としても、発達過程で必要な基本的教育内容とは何か、人に残すべき能力とは何かを検討することも必要である。(人にとって本質的な能力や人にしかできない能力の育成)

4.5 社会的論点

安全安心な社会の実現や、少子高齢化が進む中でも少ない労働力で高い生産性が上がること、個人に最適化した支援機能による多様な人の社会参加を推進することなど、人工知能技術が持続可能社会 Society 5.0 の実現に貢献する可能性は大変大きい。ただし、他の多くの道具や技術と同様に人工知能技術も1つの技術であり、その使用が社会的に強制されるものではない。個人の信条に基づく人工知能技術との関わり方の自由を確保し、人工知能技術の利用を強制されないことおよび自由に利用できることが大切である。人工知能技術は情報技術の一部として働き、外からはそれが含まれているかどうか分かりにくいことから、人工知能技術の使用の有無について明示するかについての検討も必要であろう。そして人工知能技術を利用する者と利用しない者との間に社会的対立が生じないような配慮も必要である。そのためにも、異なるビジョンや考え方を持つ者同士が対話する場をつくり、専門家の意見も含めて継続的に検討を続け、人工知能技術に対する多様な価値観を尊重すると同時に、共通基盤的に社会的に共有できる価値とは何かを考えていくことが大切である。(人工知能技術との関わり方の自由と共有可能な価値についての対話)

人工知能技術の便益を最大限に享受するには、人工知能技術に関するリテラシーに加えて、個人情報保護に関するデータの知識、デジタル機器に関するリテラシーなどがあることが望ましい。ただし全ての人がこれらを有することは現実には難しく、いわゆる人工知能技術デバインドが出現する可能性がある。例えば、本来移動弱者に便益となるべきライドシェアが知識不足により利用できず、代わりに利用する通常のタクシーがなくなってしまうたり、高額になってしまうような事態が考えられる。したがって、リテラシーや知識、資産の有無によって新たな格差や社会的コスト負担の不均衡が生じないような配慮・政策が必要である。人工知能技術によるプロファイリングなどの推定結果に基づく差別が生じる可能性について検討し対処する必要がある(人工知能技術によるデバインド、社会的コストの不均衡、差別への対処)

人工知能技術が社会で活用される場面が多くなるため、将来的には人工知能技術に対する依存や過信・過剰な拒絶など新たな社会問題や社会的病理を生じる可能性もある。人工知能技術による情報の推薦・個人最適化によって、人が触れる情報の範囲が狭くなり、またそれを普遍的なものとして受

け止めてしまう傾向が強くなる可能性もある。正しい情報の公開、議論の場の提供、教育施策などを通じた対処も必要であろう。(新たな社会的病理の可能性、対立、依存への対処)

4.6 研究開発的論点

人工知能を研究開発する者は、高い倫理観を持って研究開発に従事し²⁴、各種学会や所属機関の倫理規定²⁵、ガイドライン等を順守して価値創出に努めること、およびそれらに関して説明責任(アカウントビリティ)を持つことが求められる。特に人工知能技術はその開発速度が速く、既にある機械やサービスの内側でソフトウェアの一部として働くことが多いことから、利用者にはそれが使われているのかどうか、どのように機能しているのかなどがわかりにくいという特性がある。それゆえ、そのような状況にどう対処して、利用者に人工知能技術の使われ方や度合いに関して適切な情報を伝えることが可能かを研究開発の側面からも検討する必要がある。(倫理観、アカウントビリティ、可視化)

利用者が人工知能技術を安心して利用できる環境を整備するために、サイバーセキュリティの強化、データやアルゴリズムの改ざん防止など安全性を追求する研究開発が必要である。特に個人情報とプライバシーの保護、それをどこまで利用可能とするかの選択を安全に可能とする技術の開発が求められる。人工知能技術が制御不可能にならない技術(制御可能性)や推論・計算の過程・論理を説明できる技術(透明性)、人と人工知能技術の制御権の切り替えをスムーズにするインタフェースなどに関する研究が必要である。(セキュリティ確保、プライバシー保護、制御可能性、透明性)

機械学習に基づく人工知能技術では、確率的に妥当な結果が得られ、それが統計的に便益をもたらす。このような仕組みが社会に受容されるために、研究開発者は多くの人にそれを正しく説明することが求められる。研究開発者や報道機関等が、社会に対して人工知能技術を語る際には、技術によって得られる便益とリスクを恣意性なく的確に表現することが大切である。人工知能技術と人間社会の関係を正しく検討し、未来社会を適切に設計・実現するために、法律や経済、心理学、社会学などの人文社会科学研究者が新しい科学技術に対する知見を身に着け自らの研究に積極的に取り入れていくことや、自然科学や工学の研究者が人文社会科学研究者と共同して研究を進める必要がある。(人工知能技術に関する適切な情報伝達と人文社会科学研究、融合研究の必要性)

現在の人工知能技術は機械学習を中心に発展しているが、基盤となる理論や技術には様々あり、今後も新しい理論が出現する可能性がある。政府は、人工知能技術の多様性を確保しつつ研究開発を促進するために、基礎研究の推進やオープンサイエンスの環境を整備することが必要である。それは人工知能技術の高度化とともに頑健性と健全性を高め、社会の多様性にも対応することになる。(人工知能技術の多様性確保と多様な社会への対応)

²⁴ 総務省「AI ネットワーク化検討会議」では、「研究開発の原則」の策定として、(1)透明性の原則、(2)利用者支援の原則、(3)制御可能性の原則、(4)セキュリティ確保の原則、(5)安全保護の原則、(6)プライバシー保護の原則、(7)倫理の原則、(8)アカウントビリティの原則を含むことを発表した(2016年4月中間報告書、同年6月報告書2016)。

²⁵ 人工知能学会は、2017年2月に「倫理指針」を公表した。今後、研究機関・大学・企業内に人工知能に関する研究倫理審査委員会(Institutional Review Board: IRB)が設置されることが予想される。

第5章 おわりに

現存する人工知能技術、あるいは近い将来実現する可能性の高い人工知能技術について、具体的な事例に基づき、倫理的論点、法的論点、経済的論点、教育的論点、社会的論点、および研究開発的論点を整理した。この懇談会では、人工知能技術の研究開発にブレーキをかけるのではなく、研究開発の速度を下げずに適切なハンドルさばきをすることで、人工知能技術と人間社会との関係をより適切なものにするにはどうすべきかを検討した。人工知能技術について懸念や検討すべき点もあるが、それらを注意深く考慮して利活用することで、大きな便益を得ることが可能となる。例えば、日本は少子高齢化が進んでいるが、人工知能技術を利用することで少ない人材でもむしろ生産性を上げることができ、人は創造的で自分が楽しいと思う仕事をするのが可能となる。また、様々な地域に住む人や高齢者を含む多様な人が人工知能技術を利用することで、従来では困難だった仕事や活動も可能となり、誰もが活躍できる社会が目指せる。政府はこのような社会を Society 5.0 と名付けその実現のための研究開発を推進している。

近年の新技术には、その開発速度の速さのみならず、ある技術が進展し安定的均衡期に達して落ち着く前に次の新しい技術が出てくるという特徴がある。したがって、ある技術に対する人間社会に関わる問題についても技術がまだ進展している途中で対処しなければならない。技術の進展と移行が世代を跨がず、同一世代の中で生じるような場合には教育や社会変革について特に迅速な対応が求められる。まさに人工知能技術はそのような局面にあり、未だその将来像や最終形態が確定しているわけではないが、この懇談会では現段階と近い将来実現される人工知能技術を対象として論点整理を行った。その結果、産業競争や雇用、教育などを配慮した政策を、未来を設計する社会科学的研究と連動させていくことが必要であると示唆された。そして、今後の技術予測を行いながら、同時に人間社会との関係についての検討を続けて行くことが大切である。

このような取り組みは国内にとどまるものではない。既にいくつかの国際発信を行ってきたが、今後も国際的に協調して「人間社会を考慮した、共存する人工知能技術に関する取り組み」を続けて行く必要がある。特に、人工知能技術と社会の関係の検討について国際的な枠組みを考えることが世界全体としての持続可能社会の実現に貢献すると思われることから、世界で共有すべき価値観は何かを検討する必要がある。

人工知能技術を利用する人も研究開発者も、政府機関や民間企業、教育関係者も、全ての人がこの懇談会で得られた論点や提言について自らのこととして受け止め、私たちの未来社会をより良いものとするために具体的な議論を続け、適切な行動を取ることを期待する。

付録（別添）

付録1 人工知能と人間社会に関する懇談会の設置・活動概要

- 懇談会設置の背景
- 構成員と関係省庁
- 活動履歴

付録2 検討の詳細

- 懇談会での議論の抜粋
- 事例の検討
- 論点の絞り込み
- 共通する論点の抽出

付録3 社会との対話

- 公開意見募集
- 集中検討ワークショップ
- 日本科学未来館ワークショップ

付録4 国際連携

- 国際科学技術関係大臣会合ワーキングランチ
- 日仏シンポジウム深層学習と人工知能
- 世界経済フォーラムYoung Global Leaders & Alumni Annual Summit
- OECD デジタル経済政策委員会TFF
- 日英科学技術協力合同委員会サイドイベント：日英ラウンドテーブル

付録5 国内外の検討動向

- 国外の検討動向
- 国内の検討動向