令和５年度　第１回　人間中心のAI社会原則会議　議事要旨

１．日 時　　**令和５年４月５日(水) 13:00-15:00**

２．場 所　　**中央合同庁舎８号館８階 特別中会議室（ハイブリッド開催）**

３．出席者※敬称略

議長　須藤　　修　　中央大学国際情報学部 教授 東京大学大学院 特任教授

構成員

安宅　和人	慶應義塾大学環境情報学部 教授
	Z ホールディングス株式会社 シニアストラテジスト
岩本　敏男	株式会社エヌ・ティ・ティ・データ 相談役
浦川　伸一	損害保険ジャパン株式会社 顧問
江間　有沙	東京大学未来ビジョン研究センター 准教授
大屋　雄裕	慶應義塾大学法学部 教授
釜萢　　敏	公益社団法人日本医師会 常任理事
木俵　　豊	情報通信研究機構経営企画部 部長
武田　晴夫	株式会社日立製作所 技師長
中川　裕志	理化学研究所革新知能統合研究センター チームリーダー
永沼　美保	日本電気株式会社デジタルトラスト推進統括部 主席プロフェッショナル
樋口　知之	中央大学理工学部 教授
平野　　晋	中央大学国際情報学部 学部長・教授
福岡真之介	西村あさひ法律事務所 パートナー弁護士
福田　剛志	日本アイ・ビー・エム株式会社 執行役員 東京基礎研究所 所長
山川　　宏	全脳アーキテクチャ・イニシアティブ 代表
吉瀬　章子	筑波大学 大学執行役員・システム情報系長 システム情報系 教授

発表者

飯田　陽一	総務省 国際戦略局 情報通信国際戦略特別交渉官
今田　俊一	ソニーグループ株式会社 AI コラボレーション・オフィス AI 倫理室 統括課長

政府出席者

高村　　信	総務省 情報流通行政局 参事官
石川　勝利	外務省 軍縮不拡散・科学部 国際科学協力室 室長
工藤　雄之	文部科学省 研究振興局 参事官（情報担当）
高江　慎一	厚生労働省 大臣官房厚生科学課 研究企画官
橘　　均憲	経済産業省 商務情報政策局 情報経済課 企画官
飯野　悠介	経済産業省 商務情報政策局 情報経済課 調整官

事務局

奈須野　太	内閣府 科学技術・イノベーション推進事務局 統括官
渡邊　昇治	内閣府 科学技術・イノベーション推進事務局長補・審議官
根本　朋生	内閣府 科学技術・イノベーション推進事務局 参事官
吉澤　達也	内閣府 科学技術・イノベーション推進事務局 上席科学技術政策フェロー

4．議題

(1)　企業における AI ガバナンスに関する取組事例について
(2)　国際的な場における AI ガバナンスに関する議論の動向について
(3)　AI ガバナンスに関する議論の方向性について
(4)　意見交換
(5)　その他

5．資料

資料 1	ソニーグループの AI 倫理活動（ソニー提出資料）
資料 2	国際的な場における AI ガバナンスに関する議論の近況（総務省提出資料）
資料 3	AI ガバナンスに関する議論の方向性について（内閣府提出資料）
資料 4	政府機関における AI の導入促進にむけた調査（内閣府提出資料）
参考資料 1	令和 4 年度 第 2 回 人間中心の AI 社会原則会議 議事要旨
参考資料 2	人間中心の AI 社会原則会議　構成員名簿

６．議事要旨

(1)　企業における AI ガバナンスに関する取組事例について

ソニーグループ株式会社より、「ソニーグループの AI 倫理活動」について説明があった。

その後の質疑応答においては、次のような言及があった。

- （ソニーとして AI 倫理の取組において課題として感じていることについて質問があり、）AI 倫理活動が AI の利活用の推進を前提としている中で、実験を含む早期の研究開発の段階においてどこまで AI ガバナンスのフレームワークを厳密に適用するかのバランスについては現在社内で議論中である。
- （社内的に ChatGPT など Chatbot を利用するための指針はあるのかという質問に対して、）社内利用の指針についてはガイドラインを前提とすることとなっているが、倫理教育の中で利活用について社内的に啓発している。個別事態については社内的に議論している。
- （社内の AI 利活用において、倫理のアセスメントのやり方について質問があり、）各段階でのアセスメントをどのように行なっていくのかを流動的な状況への対応を含めて社内的に議論検討中である。

(2)　国際的な場における AI ガバナンスに関する議論の動向について

総務省より、「国際的な場における AI ガバナンスに関する議論の状況」について説明があった。

その後の質疑応答においては、次のような言及があった。

- （AI 条約のドラフトの説明における「リスクベースアプローチ」という言葉意味については、事前規制が念頭なのかとの確認があった。また、米国（AI の仕組み自体）と EU（利用されるデータ、その保護）で異なる側面のリスクに関心があり、これに対して日本はどのように対応していくか考えていった方がいいとのコメントがあった。これに対し、）EU の AI 規制法案との整合を気にしており、データ保護の側面も強く出ている。日本としての姿勢もご指導をいただきながら考えていきたい。
- （G7 関連会合での政府の立場について。リスクは AI を何にどう使うか、すなわちユースケースによって決まるため、リスクベースのアプローチが、技術全般が規制される方向性での議論とならないような主張（例えば、ユースケースに基づいたリスクベースのアプローチ）をお願いしたいとの発言があった。これに対し、）OECD の AI システムの評価の議論でもコンテキストの中で評価することが議論されているので、その方向性で AI 条約の議論も進むと考えており、G7 でも同様の方向性での議論が進んでいるとの説明があった。

また、議長より英国の視察の結果、米国との協力で生成系 AI を積極的に利用していく方向性が感じられたといった紹介があった。

(3)　AI ガバナンスに関する議論の方向性について　及び　(4)　意見交換

事務局より「AI ガバナンスに関する議論の方向性について」の説明があり、人間中心の AI 社会原則会議で議論すべき論点、議論を行う上で留意すべきことや、これから我が国が取り組んでいくべき事項等について、次のような発言を含め、全体的な意見交換が行われた。

なお、本議事要旨では、近年の大規模な深層学習により、実用レベルの言語や画像などを生成できるようになったAI技術を、大規模データ生成AI(生成系AI、生成型AI)と呼ぶこととする。

- 日本の主張は「リスクへの対応とイノベーション促進との両立」よりも「イノベーションを決して阻害しないことを前提としたリスクへの対応」という、より積極的な表現にしてはどうか。国際ルールを作ってしっかり守る国々が、それに加わらず守らない国に対してどんどん国力を落としてしまうことは絶対に避けるべき。
- その上でリスクへの対応は、リスクの網羅的なリストアップを関係国でクローズドに行い、徹底的に議論するのが良いのではないか。他国からの気づきも多々あるはず。
- 政府がリスクの規制の前にやるべきことは、関係省庁の官僚の皆様が、最新AI（次の報告にある自動翻訳やChatbotのレベルを越えて）を自ら政策検討に活用し、その実体験からリスクを実感することが重要ではないか。
- （西側諸国では）許容できないリスクが存在しているという認識は共有されている。具体的には回復困難な生命に対して生ずるリスクについては明確な規制を早い段階で講じることについては日本も受け入れるべきであるが、このような規制が過剰にならないように訴えるべき。
- AI利活用がSDGsの実現やグローバルな格差の縮小など社会全体の効率向上に利するという価値観を共有するという点を主張したら良いのではないか。
- 国家法的な法規制（ハードロー）は、状況に応じた修正などが容易ではなく、ソフトなものの中で自主規制や共同規制を排除するべきではない。
- 分野ごとに国家による法規制以外の手法（自主規制をガバナンスする）を効果的に活用すると提案したら良い。
- まずは、既存の法律・実定法を遵守して現在のAIの開発、利活用が行われているかを確認すべき。それぞれの分野において関係省庁が確認し、遵守させるべきである。「OECD・AI原則」も含む関係ガイドライン等のソフトローも、既存の実定法も、双方共に守らせることが必要である。
- 企業視点ではAI技術の進展によりその実態把握が困難とある中、自主規制を策定することが難しくなっているという実態もある。
- 新しいAI技術、システムを実際に利用して問題点を確認することで企業は自主規制としてリスク低減の実効性を示していくことが現実的となっている。この確認にはコストがかかる。
- AI利用によって生じる利用者間の格差に対して、1000人規模の調査では50%以上がその格差を許容している。
- AIガバナンス制定のための国際協力をしていくべき。
- 生成系AIなど最近の状況を踏まえて、日本の立場を変えるべきかどうか皆様の意見を聞きたい。
- 生成系AIが発展している中、知的財産の視点での議論が人間中心のAI社会原則会議において十分行われていないのではないか。
- 欧米では基本的にAIを人がコントロールできると考えているが、日本では自然への捉え方と同様で制御できず、「なるようになる」という考え方である。
- 公開書簡で述べられている巨大AI(実験)の一時停止の実現性は未定であり、AIの人による制御不能な状況の想定に対する過剰な反応について懸念がある。

- 生成系 AI の技術について理解できていないことが多いため、それに起因するリスクの予測不能性などについて議論が必要である。
- AI を含むシステムの（ELSI の）リスクのうち、法的なリスクは顕在化しているが、倫理的・社会的リスク（レピュテーションリスク）への対応が明確ではなく、難しく重要な課題である。
- AI 品質保証は従来の技術と異なる。動かしている間に品質が変わる。チェックし続けないといけない課題がある。
- AI のリスクに対応する専門的な組織が必要で、どう作っていくかが課題。
- 利用者に AI の使用の有無を明示してきた。例えばデータソースの明示など。総論ベースでリスクの議論をすると、空中戦になる。利用シーン・レベルなどのボーダーを整理しながら議論を進めることが重要。表示された回答が、確実なデータベース由来のデータか、インターネット上のデータをまとめたものか、わからない。
- 対象の曖昧さという観点から利用パターン等、情報システムと AI システムの区別や利用者に対するＸＡＩということを念頭に置き、具体的な論点の整理が必要ではないか。
- あいまいな点について、基盤モデルは文脈依存性がある。文脈によって言っていることが変わる。コントロールできない。どう対応するか求められている。
- ChatGPT はどこのだれかわからない物知りのおじさんに聞いているような印象（関西風には「何々です。知らんけど」と言われている感じ)。正しいことの中に嘘が混ざっている。知識を持った人が使えば有用。どのような専門性を持った ChatGPT なのかが重要視されるのではないか。コントロールするところがトリガーなのではないか。
- 出力を逆向きに検証して、情報のソースを特定する AI も出てくることが想定される。大元のデータの正常性の確保についての議論が重要ではないか。
- イノベーションを阻害すべきではないが、悪意を持つ場合は制限すべき。学習データの品質について留意することを日本のガバナンスとして言っておかないといけない。
- 基盤モデルはウェブデータを大量に学習することになるが、ウェブデータを使うと、個人情報が紛れていて、取り除けない。データの利活用の観点で個人情報保護法の見直しが必要かも知れない。生成系 AI に関する新しい時代の日本の AI ガバナンスにデータの観点を入れていく必要があるのではないか。
- GPT 系や、ディフュージョン系の AI が使えること、基本的理解、リスクを知ることは基本的人権である。使わないと理解が深まらないので、この点を担保しないと利用者間の分断リスクが高まることが懸念される。
- 得ているデータには Trustability と社会的な Acceptability の視点でグレーゾーンのデータが含まれていて、それが時代とともに変化する。そのため、悪意を持った間違った学習(テロ集団、敵国による操作のようなもの)を防ぐことが必要である。
- 特定のプラットフォームが圧倒的優位になり利用料金が跳ね上がることがないように複数の代替が常に存在することを担保できる社会が必要。
- 悪意を持った意志を埋め込むようなこと(悪意の埋め込み)が現実的には可能であり、これを明示的に規制して世界的な合意をする必要がある。ロボット三原則的な何かが必要である。
- AI-AIの活動が増えるので、広域でAIシステムがダウンするリスクの回避を担保するための仕組み

が必要。

- AIによる産物の知的所有権について整理、問題解決することが大事。
- AIによる生成物に対してマーケティングのために埋め込まれた広告情報等に関する明示を行う仕組みが必要である。
- 特定のクラウド基盤上への集中という社会的リスク(停電と同様の)は深刻な問題である。
- データの言語依存性リスクがあり、日本はこの点で不利であり、そのリスク低減について懸念される。
- 「これまでハードロー規制はすべきではないとの議論であったが、犯罪的な利用に関して、ハードロー規制は必要だと思うか。」という質問に対して、「悪意の埋め込みの点と AI による生成物に対するマーケティングの点はハードロー的に近いものが必要ではないか。その他は難しいのではないか。」との回答があった。
- 医療の分野、AI 成果利用環境確保が重要。不可逆的な深刻なリスクはその通り。AI の導入時点においてリスクを十分把握できるのか、わかった時点で方向転換できるかよくわからない。明確になった時点でどういう対応が可能なのか方向を示してもらえると安心できる。
- ChatGPT や GPT-4 は、おかしな動きをしないように、大人数の専門家を使ってチェックをしている。このように人手に依存しているということを認識しておく必要がある。そのうえで、使い方を考えていくべき。
- ChatGPT を使うと、企業情報が外部に流れてしまう可能性がある。なお、仕組みはシンプルなので、簡単に作れる。企業内向け GPT と公共 GPT をうまく関連させて（企業内情報をできるだけ漏らさずに、公共 GPT と連動させつつ）使う方法を企業の方には作り出してほしい。
- 情報が急速に拡散するので、顕在化した時点でリスクをマネージメントすることは困難であり、分野ごとに生成型 AI の使い方を議論する必要がある。
- G7 があり、大変動が起きつつある。各省庁、与党でも AI の検討が進んでいる。

また、次の事項について紹介が有った。

- 東京大学未来ビジョン研究センターでは、2023 年 3 月に G7 に向けた政策提言（日：https://ifi.u-tokyo.ac.jp/news/15309/；英：https://ifi.u-tokyo.ac.jp/en/news/11267/）を出している。また、提言作成時の多くのコメントから、国際協調をしていくうえでの日本の役割が期待されていることを感じた。
- AGI が重要な権力基盤を掌握する可能性（パーソナルアシスタントとして配備された AGI は、人間のユーザーを感情的に操り、偏った情報を提供し、大企業やその他の影響力のある組織を事実上支配するまでに、ますます重要なタスクや意思決定（より高度な AGI の設計と実装を含む）の責任を委ねられる可能性）についての文献（The alignment problem from a deep learning perspective）が出されている。

(5) その他

事務局より、資料 4 に基づき、今年度取り組んだ政府機関における AI 導入促進に向けた調査について、説明があった。

これに関する質問等はメール等にて事務局へ問い合わせることとなった。

以上