人工知能戦略専門調査会　制度の在り方検討ワーキンググループ（第１回）
議事要旨

１　日 時

令和８年 10 月 28 日（水）15:00～17:00

２　場 所

中央合同庁舎８号館　816 会議室

３　出席者

(1) 制度の在り方検討ワーキンググループ 構成員 ※敬称略

宍戸　常寿	東京大学大学院法学政治学研究科 教授 （座長）
生貝　直人	一橋大学大学院法学研究科　教授
岡崎　直観	東京科学大学情報理工学院　教授
岡田　淳	森・濱田松本法律事務所外国法共同事業　弁護士
岡田　陽介	株式会社 ABEJA 代表取締役ＣＥＯ ／日本ディープラーニング協会 副理事長
殿村　桂司	長島・大野・常松法律事務所　弁護士
永沼　美保	一般社団法人日本経済団体連合会デジタルエコノミー推進委員会 国際戦略ＷＧ主査／日本電気株式会社 主席プロフェッショナル

(2) 政府側参加者

塩崎　彰久	内閣府副大臣
福永　哲郎	内閣府統括官
古賀　康之	内閣府参事官
伊藤　建	内閣府参事官
羽深　宏樹	内閣府上席フェロー

(3) その他

村上　明子	ＡＩセーフティ・インスティテュート（ＡＩＳＩ）所長

４　議事概要

(1) 塩崎内閣府副大臣冒頭挨拶

議論に先立ち、塩崎内閣府副大臣より以下の挨拶があった。

- ＡＩが想定を超えたスピードで進化している。経済活動におけるＡＩの活用を見ない日はない一方で、近時のサイバーセキュリティの事案に見られるように、先端的なＡＩが生活を脅かしているとの懸念が国民の間から出てきている。ＡＩが

人間の想定を超えて、暴走してしまうのではないかという国際的な議論も深化している。どのようにイノベーションと規制を両立するのかが問題。

- 先般開催された第１回ＡＩ・デジタル推進会議ではＡＩの急速な進展に伴うリスクを最小化していく必要が示された。それを踏まえ、本ワーキンググループ（ＷＧ）においては、諸外国の制度や技術動向をみながら、日本のＡＩガバナンスについて、ＡＩ法を含む全体的な法制について見直しを進めていくこととしている。今日の情勢は、規制を強化するのか、緩和するのかという平面的な議論だけでは対応できなくなってきている。どのように安心してＡＩを使用し、挑戦していくのか、どのように国が責任をもつのか、これを考えるうえで、どのようにＡＩの安全性を確かめるのかが重要なテーマとなる。
- 新時代に新しい方法でルールを設計していくために、委員の皆様には有用な意見を頂戴したい。私も本ＷＧには、ハンズオンで参加したいと考えている。

⑵ 事務局説明「本ＷＧ開催に係る背景と目的」【資料１】

事務局から、本ワーキンググループ設置の経緯及び趣旨について、資料１に基づき説明があった。

⑶ 事務局説明「ＡＩのリスクを巡る情勢について」【資料２】

事務局から、ＡＩのリスクを巡る情勢について、資料２に基づき説明があった。

⑷ 事務局説明「ＡＩのリスク対応に係る現行の政府の取組について」【資料３】

事務局から、現行の政府の取組について、資料３に基づき説明があった。

⑸ ヒアリング「ＡＩ技術の安全性とＡＩＳＩの取組について」【資料４】

ＡＩセーフティ・インスティテュート（ＡＩＳＩ）の取組について、ＡＩＳＩ村上所長から、資料４に基づき説明があった。

⑹ 討論【資料５】

上記の発表等を踏まえ、事務局から提示された以下の議題について、参加者より意見が述べられた。

○　ＡＩのさらなる高性能化等により、ＡＩのリスクがより複雑化・深刻化しつつあると懸念されている。ＡＩ法成立以降の状況変化も踏まえ、ＡＩのリスクをどのように評価すべきか。再整理が必要ではないか。

○　現行の制度的枠組みにより、今後もそれらのリスクに適切に対応していくことはできるか。特に、どのようなリスクについて、懸念があるか。また、それらのリス

クに今後とも適切に対応していくため、ハードロー、ソフトロー含め、どのような制度的対応が適切か。

ア　リスクの再整理について

- リスクの整理については、資料２の４ページの枠組みは念頭に置く価値がある。少なくともミスユース（悪用）とマルファンクション（予期しない挙動）の両面からしっかりと対応することを優先しつつ、性的ディープフェイクについては、欧州では禁止されているし、青少年保護は社会的リスクの中でも優先順位が高い。どこまでリソースを割くかはさておき、変わり続ける環境にどのように対応するかが重要。
- ハードローで規制する場合には、対象事業者の範囲が論点となる。一般論として、ＡＩによるリスクには様々な事業者が対応すべきことであると理解している。ＡＩエージェント時代の社内リスクという意味でもどういうデータにアクセスさせるのか、どこまでの権限を付与するのかについてもっと感度を持って対応すべきであるだろう。しかし、個人情報保護の規律等で対応できる部分など、ＡＩ以外の各種の業法的な部分も含めた個別分野の規律と、ＡＩ法制の守備範囲は切り分けるべきだろう。なんでもＡＩ法制で対応すべきリスクとして取り上げるべきではない。相応の義務を課すのであれば、おのずから範囲は限られる。
- 現行のＡＩ法は、可能な限り既存の法令に基づき、様々なリスク対応するスタンスであると思うが、このスタンス自体を変える必要はない。しかし、既存法で対応できないリスクが顕在化しているのも実態。その一つの側面としてモデルの開発段階や実験段階についてもこういう問題が生じている。業法は利用に近いところで規律をしているので、開発段階等における規律がないところは、ＡＩ法制の中で規制を強めるべきではないか。その他のリスクとして、資料２の４ページで説明されていたが、悪用の増加は顕在化しているが、既存の法令の執行の強化も含め既存法令で対応できる部分も多いだろう。もう一つのマルファンクション（予期しない挙動）、特に制御喪失は今後さらに発生するだろう。これは開発段階と利用段階のそれぞれで整理すべきことがあるだろう。開発段階はＡＩ法で対応し、それ以外の部分ではAI ガバナンス等を通じて利用者において注意をすることになるのだろう。
- 大前提として、国民も含め、法律やルールの部分で方針を打ち出していくにあたり影響が大きい産業界へのメッセージングは重要。改正の理由や背景のみならず、どのようなメリットがあり、どのような社会を目指すのかについてメッセージをしっかりと打ち出すべき。事業者は、深刻なリスクに対応することで信頼を確保することには協力をすべきと考えているが、深刻なリスクが何で、なぜそのようなリスクが生じているのかについては、現状、共通理解がない。産業界の視点からは、リスクから日本を守るだけでなく、安心して投資を進められることも

重要である。まずリスクの議論と対応する規律の対象を分けて具体化すべき。専門調査会でも今起きているサイバー攻撃と将来のAIによる自律的な攻撃は全く異なるという話があったとおり、今なにかが起こっていることをベースに考えるのではなく、システム全体の検証や制御の議論をするのが良い。また、社内で事務作業に使うＡＩと、外部に接続されて契約・決済などを行うＡＩはリスクも異なる。モデル自体の能力や利用される場面、想定される被害の重大性から整理が必要。

- ＡＩの進歩に伴って生じるリスクと安全性の評価をどのように確立すべきかが重要。従来のＡＩは一問一答やそれが連続した会話となる形式で、そのなかからどのようなリスクが生じるかという議論が中心であった。ＡＩモデルの性能向上に伴い普及が拡大しているＡＩエージェントは応答に至るまでの途中の試行錯誤における想定しない動きがある。これに加えてエージェントの汎用性の高さや能力の高さが悪用されている。日本でも国産ＡＩの開発が進んでいるが、高性能なものを開発する方向性と能力はそこそこで安心して使えるものを使うという二つの方向性がある。すべてのＡＩをひとくくりにせず、能力に応じて透明性を要求することが必要。計算量に基づくカテゴリわけは必要だが、計算量は今後下がっていくと予想されるので、能力の測定による区別の基準も必要であろう。また、ディープフェイクのような、現在顕著に観測されているリスクと制御喪失のようなリスクとでは対応策は異なる。
- 技術の進化が極めて速く、SFのような将来像も視野に入れ、そこから逆算してリスクを再整理する必要がある。将来の予測は難しいので、想定外のリスクが生じること自体を想定し、アジャイルガバナンスの考え方で制度を設計することが重要。ハードロー、ソフトロー、ガイドライン、自主的な規制などを使い分けるに当たっては、リスクが顕在化した場合の影響の大きさや発生の可能性などの共通の評価軸により、対応の優先順位を整理すべき。個々のリスクに対して個別の規則を積み重ねると、制度設計上も事業者の対応上も、実際のリスクとコストのバランスに見合わない負担が生じることを懸念している。限られた政策資源や民間のリソースをどこに投入すれば、社会全体のリスクを最も効果的に低減できるかという視点が重要であり、企業が安全性を確保しながらAIの活用を継続できる仕組みが必要。
- ここでの議論や制度的な措置を考えるにあたっては、連続性を持つ問題群の中をどのように切り分けてリソースを投入し、どこについてどのような処理をするのかが重要。広い意味でのガバナンス（利活用のためのリスクを最小化する）という方向と重大な危険の議論が連続しているが、どこまでＡＩＳＩが担保するのか等を意識しながら議論すべき。また、義務付けの対象となるＡＩの範囲と、ＡＩの研究開発段階なのか社会実装なのか利用場面なのか、リスクが顕在化し損害

が生じた場合なのかの段階の切り分けが必要である。リスクの分類については多様な分類があるだろうが、ミスユース（悪用）、マルファンクション（予期しない挙動）等という分類は一つの出発点になる。ミスユース（悪用）の場合は、リスクはかなりの部分が利用者によって生じている。これに対しマルファンクション（予期しない挙動）は開発段階で対応する必要がある。基本的なＡＩ法の枠組み、各事業法や政策対応を整理してその相互作用を担保していく必要がある。この枠組み全体を通じて何を実現しようとしているのかを改めて明確にすべきである。これを明確化することでほかの国との対話や日本の特徴もでてくるのだろう。リスクへの対応といったときには専らＡＩガバナンスの話になるが、特に、利用段階でのミスユース（悪用）についてどのように考えるかは重要である。これをこの会議で議論するのかほかの会議体で議論するのかは検討の余地があるが、利用者がミスユース（悪用）をしないように自らガバナンスをする必要があると同時に、利用者の総体に一定の役割を果たしてもらう必要があるのではないか。

イ　特に懸念されるリスクへの制度的対応について

- ＥＵ ＡＩ法 55 条のような枠組みや米国州法の枠組みを参考に、事業者にリスク管理をしっかりさせて、情報の収集と報告徴収ができるようにすることは不可欠だろう。加えて、透明性については米国州法のようなフレームワークに従っていくのだろう。プラットフォーム系の検討会ではよく話題になるが、リスクに対する対処状況は日本に一定の制度的なものがないと日本の文脈に即した公表をすることが事業者としてもできない。また、ＡＩＳＩを規制全体の枠組みでどのように位置づけるべきかが問題。
- ＡＩ等の重大な脅威が急迫している中で、分野別に既存の法令やガイドラインをベースに発展させて対応する路線を崩す必要はないが、悪用された場合や予期せぬ挙動が生じた場合、非常に速いタイムラインで計り知れない損害が生じる可能性があるので現行のソフトローを中心とした制度で今後も続けていくべきかについては、再考の余地がある。また、ＡＩは海外の事業者が開発するものも多く、海外事業者への対応としてのサンクションの裏付けを持った規制をしていくべきではないかという問題意識とは、向き合わなければならない。
- 実効性の確保は重要である。従来の法律にあるような指導、勧告、命令という段階的アプローチを原則とすることは合理的であるが、国民の安全等に問題が生じる場合には、罰則以外にも、速やかに緊急命令ができるような観点は検討するべきであろう。罰則については直罰の是非も論点になり得るかもしれないが、いずれにせよ事後的な処罰の問題とは別に、差し迫った危険を速やかに防止する措置の在り方もきちんと考える必要がある。

- 特に開発者向けの規律は現行のＡＩ法では十分ではない。ＡＩ法第16条に基づく任意の情報収集の枠組みは機能しているところとそうでないところがある。一定の情報開示についてはハードローでやるべき段階にあるかもしれない。AIの進展が早いため、具体的な基準はガイドラインに落としていくことになるだろう。規制強化の一方でイノベーションへの配慮は必要で、ハードローの適用を受ける事業者への配慮も重要だろう。一定の情報を開示させるとして、秘密も保護する必要がある。自主的に開示している内容も加味しながら、過度な負担にならないように配慮すべきだろう。また、海外の事業者への実効性を持った執行ができるのかは注目されるところであろうと思われ、慎重な議論が必要。
- 制度を検討するにあたっては、リスクの議論と対応する規律の対象を分けて具体化すべき。法的義務を設ける場合、対象となる事業者や行為、求める措置を明確にする必要がある。企業は自ら対象かどうかが判断できないと、やはり萎縮の方に走る。また、ＡＩサプライチェーン全体で責任の分担が必要。海外企業などの他社が開発したＡＩを組み込む場合、利用企業が基盤モデルの内部や学習の過程までを遡って把握して安全性を保証することには限界がある。役割を整理の上で下流の事業者に必要な情報が流れるようにしてほしい。加えて、実効性と予見可能性の両立を目指すことが重要である。報告する情報の範囲や利用目的、営業秘密、企業情報の保護を明確にして協力しやすい仕組みを整備しつつ、既存の報告制度との重複は避けるべき。報告のインセンティブ設計の議論も重要。具体的なところはガイドラインで定めるとしても、ガイドラインも乱立しており、事業者の負担になっている実態もある。日本の中での安定した事業環境と投資促進ための環境整備が必要で、そのためには、海外事業者への実効性がなく国内事業者ばかりに負担がかからないようにすることも重要。
- ＡＩの安全性評価について様々言及されてきたが、サイバー空間などでＡＩが悪用されることへの防止措置についても研究開発を進めることが必要。ＡＩのリスクから国民を守るための研究や体制作りが重要。ＡＩＳＩの機能も強化していかなければならない。
- 情勢も急速に変化し、あらかじめリスクを明文化しきれない部分については、社外取締役や倫理委員会等のコーポレートガバナンスを活用することも考えられる。例えば、ハイリスクなＡＩを提供する企業に対し、ＡＩに関する特別な役割を持つ取締役の選任を義務付けるなども考えられる。海外事業者に対しても、日本国内でハイリスクなＡＩを提供する場合に必要なガバナンス体制を求めることは、けん制としても重要ではないか。制度の設計によっては、日本で開発されたＡＩを使用することの利点にも繋がり得る。さらに、このような制度や仕組みをグローバルサウスをはじめとする海外に展開し、日本の制度・政策との接続を図

ることで、日本のＡＩ事業者が海外で活躍しやすくなる土台を作ることも考えられる。

- ハードローで問題になるのは、特に最先端のＡＩの研究開発が行われているモデルが高速に更新され想定しないリスクやマルファンクション（予期しない挙動）が生じる状況についてハードローに基づいて何をするのが妥当かという点。特に、それが広汎であったり、不明確でないかについては集中的に議論が必要。
- 現在のＡＩ法第 16 条においても権利利益の侵害が生じた場合の検討とはっきりと書いているが、この場合の情報提供の求めは制裁的なものではないというのが現行法。権利利益の侵害のおそれがある場合に報告を求めることができるのか、強さとして報告徴求をすることができるのかが問題。外国のフロンティアＡＩラボを持っている企業には任意で対応してほしいものの、協定に近いようなもので情報を出してもらえるうちはともかく、信頼の原則を裏切るような行為や、もともとＭＯＵを結んでいない国の企業に対しては報告徴求できるようにしておく方が、ＭＯＵを結ぶことが有利というインセンティブが生じるかもしれない。このような枠組みも検討していく必要があるだろう。このように得られた情報を全体として的確に収集分析され、国民に共有されることが重要である。ＮＣＯでは官民連携を大きな柱としているが、協議会のような枠組みは参考になるだろう。また、このような方法で報告を求める場合に、その報告結果をどう使い誰に共有できるのかという点も重要。ＡＩＳＩに機微な情報も含め、情報共有ができるのかという検討を行う必要がある。
- ＡＩの分野の報告枠組みはサイバー分野の報告の枠組みとも似たものになると想定されるが、企業内でも分野によって担当者や専門性が異なることを踏まえ、サイバーセキュリティとＡＩの安全性、セキュリティの間で運用の連携・整合性に配慮してシンプル化を意識していただきたい。
- 基本的には開発者の顔が見える前提で議論が展開しているように思われるが、昨今、オープンウェイトモデルの悪用も懸念され、誰が作ったのかわからないモデルが公開され、使われていることもある。このようなものに制度として何ができるのかは論点であろう。
- ミスユース（悪用）にも２種類あり、確固たる意思がなく悪質なＡＩの利用が行われるケースをけん制するためにはガバナンス等の方向性があるのだろうが、犯罪を計画的に実行しようとしている人をやめさせるのは困難である。これについてどのように扱うのかは検討せざるを得ないだろう。これは本WGの範囲を超えるのかもしれないが、ルールを守らない人にどう対応すべきかは重要である。ＡＩの安全性を確認する際には、想定される使い方をしない悪用をする人が存在することを前提に対応を考えることが安全性の考え方の中で重要だろう。

- 安全性を考えるうえでは、事業者から情報収集をするにあたっての政府の権限又は事業者の義務について、現状以上の法的な規律を設ける必要があるか。そもそも情報収集を制度の軸にするのか。また、仮にハードローで規律する場合には、ＡＩがリリースされて日本で利用される段階で規律するのか、開発段階から規律を及ぼしていくべきか。さらに、捕捉するのは、事業者単位なのかモデル単位なのか。モデル単位だとしたときに、性能を測る単位はFLOPS数でいいのか、そのようなパラダイムは変わっているのか。これらは、今後の検討を進める上で、重要な論点ではないか。

(7) 福永内閣府統括官閉会挨拶

閉会に際して、福永統括官より、引き続き議論を重ねていってほしい旨の挨拶があった。

以　上