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２回）
議事要旨

１　日 時

令和８年 11 月 11 日（水）10:00～12:00

２　場 所

赤坂グリーンクロス　22 階会議室（オンライン併用）

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、以下同じ

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
恒藤　晃	内閣府審議官
伊藤　建	内閣府参事官
輪倉　真也	内閣府企画官
吉原　博紀	内閣府企画官
津田　麻紀子	内閣府政策参与
羽深　宏樹	内閣府上席フェロー

(3) ヒアリング発表者（発表順）

根本　宗記	ＮＴＴ株式会社技術企画部門 AI ガバナンス室室長
城　譲	Sakana ＡＩ株式会社政策渉外責任者
井尻　善久	ＳＢ Intuitions 株式会社代表取締役社長兼 CEO
吉田　朗	Anthropic Japan 合同会社公共政策責任者

Eunice　Huang　Google 社 APAC Head of AI & Emerging Tech Policy
斉藤　理　　Google 社シニアマネージャー
大久保　和也　OpenAI Japan 合同会社 Head of Policy and Partnership

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とおり挨拶があった。

- ＡＩは社会の様々な問題や経済の中心に位置づけられており、ＡＩの可能性とリスクについて、聞かない日はないくらいである。
- 第１回では委員から、ＡＩのリスクと今後必要となる制度的対応について忌憚のない意見をいただいた。第二回となる本日の会合では、前半はＡＩを開発提供する事業者から、サイバー攻撃を含むＡＩの様々なリスクに対する考え、各社の安全性確保の取組、制度設計に関する率直な意見を伺う。事業の現場から見える忌憚のない声をいただきたいと考えている。後半は有識者の討論に移り、制度的対応が必要なリスクの範囲、規律の対象とする事業者、制度の具体的な内容について、一歩踏み込んだ議論をお願いしたい。

(2) 事業者ヒアリング【資料１－２】

事務局から、事前に提示されていた以下のヒアリング事項について説明があった。

- サイバー攻撃やＣＢＲＮＥテロ（※）等の高度なリスクを含むＡＩの様々なリスクについて、ＡＩを開発する立場からどのように考えているか
- 自社における現行のＡＩ安全性、リスク対応の取組
- ＡＩの安全性について、どのような制度が必要だと考えるか。ソフトロー、ハードロー含め、制度設計に関する要望

その後、

- ＮＴＴ株式会社
- Sakana ＡＩ株式会社
- ＳＢ Intuitions 株式会社
- Anthropic Japan 合同会社
- Google 社
- OpenAI Japan 合同会社

より、ヒアリング事項についての発表があった。また、委員から事業者に対して、発表内容に関する質問があった。

※　Ｃ（Chemical：化学）Ｂ（Biological：生物）Ｒ（Radiological：放射性）Ｎ（Nuclear：核）Ｅ（Explosive：爆発性）物質を使用したテロの総称

(3) 事務局説明「前回の議論のまとめ」【資料１－３】

前回の議論における有識者の意見と、それを踏まえた今後の方向性の案について、事務局より説明があった。

(4) 討論【資料１－４】

事業者ヒアリング等を踏まえ、事務局から提示された以下の論点について、委員より意見が述べられた。

論点１　現行のＡＩ関連制度を踏まえ、更なる対応が必要なリスクは何か。また、一定以上の性能を有するＡＩを規律する場合、その外延をどのように設定すべきか

論点２　ハードローで対応する場合、どのような者を規律の対象とすべきか（事業者か利用者か、事業者のうちでも、どういった事業者を対象とすべきか）

論点３　ハードローで対応する場合、どのような義務を課し、国にどのような権限を設けるべきか。また、ＡＩＳＩはどのような役割を担うべきか

論点４　上記以外について、既存法令やガイドラインで、どのような追加的な措置を講じていくべきか（どのような事業者に、どのようなリスクに対応すべきか）

論点５　その他、制度設計にあたり、どのような点に留意が必要か

● 資料１－３で示された大枠の方向性には賛成。特定のリスク領域は、①医療・看護・介護など人の生命・身体に直結する領域、②金融（アルゴトレード等）など大きな経済的影響を及ぼす領域、③これまで高度な専門性を持つ人がリスク管理してきた企業・産業の根幹業務・システム、の三つの観点が重要。加えて、フィジカルAI領域についても、一度物理世界に損害を与えると元に戻せない点を踏まえ、議論する必要がある。

一定以上の性能を有するＡＩの線引きについては、計算量の一律基準もあり得るが、Transformer中心のモデルが今後も続くかは誰にもわからず、何を認め何を禁止するかをハードローで固定的に定めると技術の進歩に追いつかなくなる。性能に加え、Chain of Thought監視なども含めた複数の防衛線の構築が重要。

海外事業者と国内事業者の不公平感をなくすため、ある程度の強制力を持たせることが重要。国の権限はリスクに対して比例原則で対応し、一律の監督も重要だが、情報収集、事故報告、是正対応の仕組みを検討することが重要。ＡＩＳＩは細かいソフトロー整備の役割が大きく、そのための人材強化が重要。飛行機のオートパイロットの仕組み・制度を参考にしながら、現実的な制度を整備していくことが重要。

現状の法制度上ＡＩの利用が難しい点が多数あるので、「この条件に則っていればＡＩを使ってよい」というホワイトリスト方式にすれば、日本の企業の特性

上、利用可能な条件が明確になることでＡＩを活用しやすくなり、世界で最もＡＩを開発・活用しやすい国という目標とも整合する。

メリハリのある制度設計が日本企業の安全なグローバル展開の後押しになる。グローバル基準と日本基準を合わせ、日本基準がデファクトスタンダードになれば、日本のＡＩ企業の海外進出の一助になる。

● 前回、目指す社会とメリットを示すことが重要と申し上げたが、深刻なリスクへの適切な対応が利用者の信頼を高め、AI の普及を後押しする「信頼と普及の好循環」を、日本のＡＩ産業全体の発展につなげることが重要。

論点１は、サイバー攻撃、化学・生物の悪用、制御喪失を含む予期しない挙動による重大被害を重視すべき。法的義務を課すなら、被害の重大性、発生可能性、現行法令や自主的取組で対応できない理由を示す必要がある。対象とする AI の要件は、将来的に少ない計算量でも高性能化が進むので、計算量だけでなく能力・影響に注目し、評価の基準は、ＡＩＳＩや国際枠組みとも整合させ予見可能なものにすべき。

論点２は、対象モデルを特定しリスクを管理できる開発・提供事業者を報告の求め先とし、組込事業者や利用企業に一律に同じ義務を負わせないよう対象者と行為を明確にすべき。海外事業者にも同等の実効性を及ぼすことは不可欠。

論点３は、日本で公開・提供される対象モデルの重大案件の報告や、任意の協力が得られない場合に求める情報提供について、目的と対象を絞った強制力ある仕組みを検討すべき。ハードローの義務は情報の収集と提供を基本とし、安全措置の義務は技術の進展が速いので、法定するより事業者自ら公表し国と情報共有し国が分析・検証する形が現実的。公表では悪用防止や秘密保護に配慮が必要。勧告・命令・罰則は要件を明確にし、必要性を検証のうえ段階的に。事故の発生自体を制裁対象にすると報告が萎縮しかえって情報が集まらないことも考えられる。報告の目的・共有先・営業秘密の保護を明確にし、海外当局への報告との重複も避けるべき。ＡＩＳＩには評価基準の作成、収集情報の分析・検証を期待する。海外で開発・試験中のＡＩの逸脱が国内に被害を及ぼす場合の情報収集について、国際連携を含めた検討が必須。

論点４は、報告対象外でも基盤モデルの悪用や事故への備えは欠かせず（英国家サイバーセキュリティセンターもＡＩとサイバー攻撃手法の連動を指摘）、現行法令やガイドラインに基づき利用権限の制限や人による確認で対応すべき。基盤モデルを組み込んで利用する下流の事業者はモデルの内部を確認できないので、安全な利用に必要な情報が上流の開発・提供者から届く仕組みが重要。既存の業法やガイドラインを活用しながら、想定される利用上の制約や安全対策を示すことも考えられる。国外で公開されたオープンウェイトモデルは規律が及ばな

いと思われるが、重要インフラ防御や現行法令を組み合わせた多層防御が重要。

論点５は、国内事業者だけに実効性が及んで負担がかからないよう、公平性、広島ＡＩプロセスとの相互運用性による報告の重複やガイドライン乱立の回避、利用目的・共有先・秘密保護の明確化が必要。高性能モデルの日本での提供が滞ると産業競争力だけでなく防御能力にも影響するので、安全確保と先端技術へのアクセスを両立し、国民と事業者が安心してＡＩの恩恵を享受できる制度を検討すべき。

● 資料１－３で示された方向性に大きく共感する。日本の法制度が及んでいない事業者に、一定のハードローとしての規律をかけていくことになるだろう。

規律の対象となるＡＩの外延は、国際的には計算量基準を使用している例が多く一つのベンチマークになるが、それに限らず能力や利用者数などを考慮して指定・特定できる柔軟な枠組みが望ましい。

ＡＩＳＩの評価能力は極めて重要。現状は独立した機関だが、国の調査、徴収した報告の検証、場合によってはモデルアクセスまで協力できるよう、法律上の位置づけをしっかりできれば望ましい。

我が国独自で取り組むべき検証は多いが、情報公開・透明性は各国制度が共通化してきている。広島ＡＩプロセスの透明性フレームワークや米国州法のフロンティアフレームワーク等に、義務が大きく重複しない限りで国内でも法的位置づけを与え、詳細が必要な場合に個別に報告徴収する、という国際枠組みに準拠する仕組みを考える余地が大きい。

● 論点１では、リスクの内容・性質の議論と、そのプロキシとしてどの程度のＡＩを対象とするかの議論は連動しつつも観点が異なる。対象リスク（サイバー、ＣＢＲＮＥ等）は委員の認識がおおむね一致し、性能も計算量を重視しつつ他の要素も見る必要がある。問題は具体的な範囲の決め方で、法令に明確に書く部分と、予見可能性を保ちつつＡＩＳＩの調査を踏まえ規制対象に組み入れる部分の制度的な落とし込みを整理すべき。

論点２では、「ＡＩ事業者」だと解像度が低いので、まず上流の研究開発・提供事業者の範囲を明確にすることが大事である一方、最先端モデルを組み込む事業者にどこまで規律を及ぼすかは難しい論点。オープンウェイトモデルそのものは規制しないにしても、改変して使う事業者・利用者への手当てについては委員やヒアリング企業から重要な指摘があった。ただ許可制は非現実的で、様々な形で利用する実態を把握し対処できる構えで検討が必要。

論点３では、ヒアリング企業各社は高度なガバナンスのフレームワークやレッドチーミング等で高い水準の取組をしており、それを前提に義務内容を検討して

いくべきだろう。加えて、情報提供義務、インシデント発生時や具体的危険の顕在化時の早期報告義務が必要で、平時のモニタリングを国側でできるようにすることが大事。

前回も指摘があったが、国の権限行使とＡＩＳＩの関わりはかなり整理が必要。公権力の発動にＡＩＳＩが密接に関わると、人や情報が集まる今の良さを損ねかねない。ＡＩＳＩの体制は維持し、公権力の発動は担当大臣か本部かで政府内部の一定の手続きを取り、事業者の異議も含め現行法令と合わせて検討すべき。

調査権限を日本特有のリスクに絞るべきというグローバル企業からの指摘は、コストパフォーマンスの面でも重要。一方で、国民の不安を払拭し信頼を高めて経済成長につなげるには、全体像を政府として評価・公表する能力が必要で、調査内容と公表を１対１で対応させず、一般への公表・注意喚起はむしろ広く構えるべき。

論点４では、現行法令やガイドラインで対応するリスクについて、関係省庁でも議論が進んでいるが、重要なのはＡＩ戦略本部に情報が集まること。所管の抜け落ちとＡＩ担当部局側からのマイクロマネジメントの両方を避け、誰にどの規律が及んでいるかを分析し、ＡＩ法・所管官庁・民間ガイドラインの役割分担を発信・助言できることが大事。

● 全体として資料１－３で示された案の方向性でよいと感じている。

ヒアリングした事業者からは国ごとに基準や義務の内容が違うと大変だという話もあったので、線引きは、国際基準に準拠したほうがよいが、どこに線を引くべきかの科学的根拠や基準が明確になっているわけではないので、ＡＩＳＩ等で柔軟に検討・設定できるようにしておくのが望ましい。

対象は基本的に事業者。悪用する利用者は現行法令で対応できる部分は対応し、漏れる部分を現行法令のアップデートで対応するのか、ＡＩ法で対応するのかを検討すべき。

ヒアリングでは既に実施している安全対策や透明性報告にどの会社もあまり差がなかったので、新たに法令で規定するのは透明性を高めるための報告と安全対策の義務でよい。事前報告か事後報告かは議論があると思う。

● 資料１－３で示された大きな方向性に異論はない。リスクは多岐にわたり現行のソフトロー・業法・個別法のアップデートも必要だが、ＡＩ法が優先的に対処すべきなのは、フロンティアＡＩモデルの開発・提供を中心としたサイバー攻撃、ＣＢＲＮＥ、インフラへの影響などになる。

ＡＩ法の領域とそれ以外はある程度区分けが必要だが、完全には切り分けられ

ず一定の重複はあり得るし、重複があるからというだけで直ちにＡＩ法で扱うべきでないということにもならない。中長期的には守備範囲が広がり得るが、最初から何でも取り込もうとして全て頓挫するのは本末転倒なので、優先度の高いところからという方向性に賛同する。

「事業者か利用者か」という整理はややミスリーディング。利用者には事業者もいるし、事業者にもモデルを開発・提供する者と、ライセンスを受けてサービスに組み込む者（ＡＩ事業者ガイドラインでいう開発者と提供者）がいるが、両者は異なる。提供者側はむしろ現行法令・業法にフィットする。解像度を上げて記載した方がよい。下流ほど現行法令やソフトローで対応しやすく、新たな法令の優先対象は上流のフロンティアモデル開発・提供者になる。

オープンウェイトモデルは今後も重要な論点で、個人的にも現時点で明確な結論を出せているわけではない。実効性の限界やイノベーションとの緊張関係がある一方、日本向けに提供される高性能オープンウェイトモデルを完全に規制から外すのが正当化されるのかは疑問も残る。規律内容に差をつけることはあり得ても、完全に外してよいかはもう少し議論してよいのではないか。

規律対象の閾値について、計算量や利用者数を複合的に見るアプローチもあり得るが、本当にリスクの高いモデルは国内利用者が少なくても必ずしも危険が少ないわけではない。利用者数要件を課すにしても閾値を高くすると実効性に問題が出るので、実効性が損なわれないよう議論が必要。

有事だけでなく平時の危険低減措置とそのモニタリングが非常に重要。営業秘密や悪用のおそれがあるので、全て一般公表という規律は合理的ではないが、少なくとも政府への定期報告を含めた報告は検討してよい。

ＡＩＳＩの役割強化には賛成。これまでの技術評価手法や基準策定の支援、評価の実施主体としての役割は高く評価できる。ただ、法的義務の遵守確認や規制の執行に直接結びつくと、従前のＡＩＳＩに求められる役割の中での民間からの協力をかえって阻害しかねない。どこまでやるかをきちんと議論して切り分けて社会に発信すべきで、なし崩し的に独立評価と規制遵守確認の区分けが曖昧になると、各ステークホルダーが動きにくくなる。

● 対象リスクは大規模なサイバー攻撃、悪用によるＣＢＲＮＥテロ、制御喪失が中心。ハードローの対象者はフロンティアモデルの開発・提供者に限定し、指定要件の線引きはモデルの性能を基本とするのがよい。

利用者規模を必須要件にすると、それに満たないものが対象外になり適切でない。考慮するなら、対象範囲ではなく義務の程度を調整する要素とすることも考えられる。計算量だけでは今後カバーできないものが出るので、ＡＩＳＩで実質的な能力評価を行い一定の場合に対象とする仕組みも有用。

事前評価できていないモデルが重大事案を起こす場合もあり得るので、事業者に意見陳述の機会は与えつつ、緊急時には一定の報告を求められるようにしておくことも考えられる。

オープンウェイトモデルも基本的には対象とすべきだが固有の難しさがあり別途の考慮が必要。開発者が分からないものは名宛人が不明なので、頒布プラットフォームとの協力についても検討が必要。安全対策機能の利用者による公開後の除去にどこまで耐えられるかという安全性評価を追加で求めることも考えられる。

義務の内容は事務局案におおむね異存はない。安全確保措置は、求める項目は下位規範等で定めつつ、具体的な内容は事業者の裁量を認める柔軟なルールが必要。

インシデント報告は事後だけでは不十分で、おそれを認識した段階で求めるべき。タイミング・フォーマットは国際的な協力関係をもとに他国と大きくずれないようにし、同等制度を持つ他国への報告内容をそのまま提出すれば重複部分は日本の義務も果たしたとみなす制度を導入することによる負担軽減も考えられる。実効性の担保として立入検査権限も考えられるところ、海外事業者は海外での立入検査のハードルが高いので国内代理人の選任も検討すべき。

資料１－４のＤの領域（既存法令やソフトローで対応を促していく分野）も非常に重要。サイバーテロやＣＢＲＮＥテロのためにＡＩを悪用する利用者がいるので、きちんと処罰できるか、業務妨害等の刑事法や化学兵器関連法令の法定刑がＡＩによる大規模犯罪を前提に適切かも含め再検討が必要。オープンウェイトモデルの安全対策機能を取り除く行為や危険なモデルを公開する行為に現行法令で対応できるかも精査が必要。

制御喪失は開発者への規律だけでは不十分で、ＡＩエージェントを実装する利用者へのルールの明確化が必要。ＡＩ事業者ガイドラインを骨格に、実装時の考慮点や問題発生時の関連事業者間の責任関係を精査すべき。

以　上