AIネットワーク社会推進会議
AIガバナンス検討会（第 28 回）
議事概要

1．日時

令和 7 年 12 月 2 日(火) 13:00～15:00

2．場所

オンライン開催

3．出席者

(1)構成員

平野座長、大屋座長代理、荒堀委員、上原委員、浦野委員、江間委員、落合委員、小俣委員、木村委員、小塚委員、財津委員、斉藤委員、佐久間委員、三部委員、実積委員、高木委員、高橋委員、瀧澤委員、田丸委員、千葉委員、豊田委員（代理出席）、中川委員、成原委員、西田委員、林委員、山川委員、山田委員、須藤 AI ネットワーク社会推進会議議長

(2)オブザーバー

内閣府科学技術・イノベーション推進事務局、内閣府知的財産戦略推進事務局、公正取引委員会経済取引局総務課デジタル市場企画調査室、個人情報保護委員会事務局、金融庁総合政策局イノベーション推進室、消費者庁消費者安全課、デジタル庁戦略・組織グループ AI 実装総括班、デジタル庁省庁業務サービスグループ政府の AI 調達・利活用ルール形成班、文部科学省研究振興局参事官（情報担当）付、文化庁著作権課、厚生労働省医政局研究開発政策課、農林水産省輸出・国際局知的財産課、経済産業省商務情報政策局情報産業課 AI 産業戦略室、経済産業省商務情報政策局情報経済課、防衛省整備計画局サイバー整備課、情報通信研究機構 AI 研究開発推進ユニット、科学技術振興機構社会技術研究開発センター企画運営室、理化学研究所数理・計算・情報科学研究推進部、産業技術総合研究所情報・人間工学領域、AIセーフティ・インスティテュート

(3)総務省

布施田国際戦略局局長、桐山国際戦略局次長、寺村国際戦略局情報通信国際戦略特別交渉官、後白国際戦略局国際戦略課 AI 政策推進室室長、藤本国際戦略局国際戦略課 AI 政策推進室課長補佐、他

4．配布資料

資料 1 今年度の活動について

資料 2 国内動向報告（AI 法等）
資料 3 国際動向報告（広島 AI プロセス等）
資料 4 AI事業者ガイドラインの更新に向けた論点
参考 1 AIネットワーク社会推進会議 AIガバナンス検討会構成員名簿
参考 2 AI事業者ガイドライン（第 1.1 版）本編
参考 3 AI事業者ガイドライン（第 1.1 版）別添

5. 議事要旨
5－1. 開会

5－2. 議事
（1）今年度の活動について
事務局より、資料 1 に基づき、AI 事業者ガイドラインの検討における実行体制や AI 事業者ガイドライン更新にあたっての検討事項、令和 7 年度の活動スケジュール案等について説明がなされた。

（2）国内動向報告（AI 法等）
まず、最上内閣府科学技術・イノベーション推進事務局参事官補佐より、資料 2 に基づき、人工知能関連技術の研究開発及び活用の推進に関する法律（AI 法）の概要や、同法令に基づく AI 基本計画骨子・適正性確保に関する指針骨子について説明がなされた。次いで、事務局より、政府における AI に関する取り組みの一部について説明がなされた。

（3）国際動向報告（広島 AI プロセス等）
寺村国際戦略局情報通信国際戦略特別交渉官より、資料 3 に基づき、広島 AI プロセス等についての説明がなされた。

（4）AI 事業者ガイドライン更新に向けた論点
事務局より、資料 4 に基づき、AI 事業者ガイドラインの更新に向けた論点について説明がなされた。

（5）意見交換
主な質疑応答等は以下のとおり。

＜議事 2 国内動向報告（AI 法等）に関して＞
【落合委員】
各種取り組みが着実に進展している状況について理解した。分野別に見ると、金融庁においては

ディスカッションペーパーが公表されているほか、医療分野においても AI 関連の議論が進められている状況にある。今回の政府調達や民事責任のように、特定の分野や具体的なテーマに即した検討は、今後さらに増加していくものと考える。こうした個別の検討と今後策定されるガイドラインとの関係性、及び相互に影響を及ぼし合う部分について、いかなる観点から整理していくべきか、事務局の見解を伺いたい。

【藤本課長補佐】
AI については、関係各省における取り組みが進展していくことから、AI 事業者ガイドラインについてもそれらとの整合を図っていく必要があると考えている。AI 事業者ガイドラインの対象者は、事業者のみならず、政府や地方公共団体等も対象としており、その範囲は広い。そのため、各主体に共通して適用される事項や、共通するリスク及びその対策を中心に記載することとし、業種固有の事項については、脚注等において当該業種に特化したガイドライン等を参照する形で整合を図るのも一案と考えている。引き続き、政府内における議論や動向については、事務局において広く把握に努め、関係各省との整合を図ってまいりたい。

【落合委員】
そのような形で進めていただければ、相互の参照関係が明確となり、今後の運用においても有益であると考える。また、個別具体的なリスク管理については、各分野において具体的な想定を踏まえて実施することが、より合理的であると考える。本会議等において全体のガイダンスを策定し、個別分野との役割分担のもとで進めていくという方針で、今後も取り組んでいくことが望ましいと考える。

【平野座長】
内閣府の人工知能戦略本部については、全閣僚が構成員となっており、その事務局側[人工知能戦略推進会議]には各省の局長級が参加している旨、城内大臣が国会答弁において言及されている。このため、同会議等において関係各省間の連絡調整が行われるものと認識している。

【中川委員】
AI については、全般的な法律のみならず、個別の業法にも関連する部分が非常に多くなるものと考える。落合委員のご発言に補足すると、AI 事業者ガイドラインにおいて掲げられている一般原則が、いずれの業法に該当するかという対応関係を整理する必要があると考える。その際、一つの業法のみならず、複数の業法にまたがる可能性もあることから、参照すべき箇所を丁寧に整理することにより、実際に業法関連の作業に当たる方々にとっても活用しやすくなるのではないかと考える。

【平野座長】
本件については、事務局とも当該点に留意しながら検討してまいりたい。業法に関しては、AI 法に

係る国会答弁においても言及があったとおり、個別の業法については各省庁が最も精通しており、各省庁が主導して詳細設計を行うこととなる。他方、AI 事業者ガイドラインは対象範囲が非常に広いことから、現在策定を進めている AI 法に基づく指針との整合性を図りつつ、各業法においていずれの事項が該当するかについても、関係各省と意見交換を行いながら検討を進めていくこととなるものと考える。

＜議事 3 国際動向報告（広島 AI プロセス等）に関して＞

【実積委員】

広島 AI プロセス・フレンズグループの参加国については、OECD 加盟国以外の国が徐々に増加してきていると認識している。他方、AI に関しては、グローバルサウスがこれまでのガバナンスに係る議論から十分に参画できていない印象を受けているところである。広島 AI プロセス・フレンズグループは、グローバルサウスに対して日本の知見や AI ガバナンスに関する貢献を発信する上で、非常に有効な機会であると考える。グローバルサウスとの関係や南北問題に特化した働きかけについて、今後総務省において何らかの計画があるか、ご教示いただきたい。

【寺村特別交渉官】

南北問題といっても、実際には発展段階が異なる様々な国が存在しており、可能な限り各国の状況に寄り添った形で、日本のそれぞれの段階での取組をモデルケースとして紹介などしつつ多様な支援を実施していくことが重要であると考えている。したがって、単に南北問題として捉えるのではなく、広島 AI プロセスという共通の概念のもとで各国が連携していくためにはどのように取り組むべきかという観点から検討を進めていくことが肝要であると認識している。

【実積委員】

広島 AI プロセスについては、各方面から伺うところによると、EU AI Act のような拘束力を有する法令ではなく、各国が自発的に参加し、共同で遵守していく枠組みが相当程度の賛同を得ているとのことである。今回の取り組みが、いわゆるグローバルサウスと呼ばれる国々にも受け入れられることを期待したい。

【寺村特別交渉官】

本件については、私自身も海外、特に東南アジア諸国の関係者と接触する機会が多く、その際、EU が規制色の強いルールを導入する中で、日本の立場はいかなるものかという質問を頻繁に受けるところである。そうした場において、我が国としては、規制的なアプローチではなく、AI 産業のイノベーションを促進することが極めて重要であると考えており、可能な限り規制によらない形での取り組みを志向している旨を説明している。この立場については、相手方から賛同を得られることが多い。こうした我が国のスタンスについて、今後より一層積極的に発信していきたいと考えている。

【小塚委員】
広島 AI プロセス・フレンズグループの理念自体は大変意義深いものであり、日本がこうした働きかけを行ったことは画期的であると評価している。他方、このプロセスのガバナンスについてお伺いしたい。当初の趣旨としては、広島 AI プロセスを G7 以外の国々に拡大することにあったと理解している。しかしながら、フレンズグループから独自の意見や要望、さらには広島 AI プロセスにおいて決定された事項に対する改定の要請等が出された場合、それらをどのように取り扱っていくのか、またどのような枠組みとなっているのか、その点についてもう少しご説明いただきたい。

【寺村特別交渉官】
まず、フレンズグループ自体は自発的な取り組みであるが、G7 との関係をどのように整理していくかという点についてご説明する。広島 AI プロセスが G7 において立ち上げられた経緯を踏まえると、これを改善していくに当たっては、G7 全参加国の了解を得る必要があることは確かである。他方、フレンズグループにおいて議論された内容については、例えば私から G7 の会議の場においてフレンズグループの取り組みを紹介するとともに、同グループから出された意見等を、大臣会合に先立つワーキンググループ等の場で共有することが考えられる。これにより、各国に問題意識を持っていただき、当該意見について議論を進めていこうという機運を醸成することが、最も有効な仕組みになるのではないかと考えている。なお、フレンズグループについては、来年 3 月に第 2 回会合の開催を予定している。同会合においては、あくまで自発的な取り組みであることが前提であり、非拘束的なアクションプランみたいなものを作れないかと考えている。ただし、詳細については未定であり、最終的にいかなる方向性となるかは現時点では明らかでないが、可能であれば方針を取りまとめ、G7 に提示したいと考えている。

【小塚委員】
結局のところ、形式上の主従関係は変更できない中で、本プロセスを提唱した日本がリエゾンとして尽力していくということであると理解した。これは大変意義深い取り組みである一方、相当な負担を伴うものと思料するので、引き続きよろしくお願い申し上げたい。なお、将来的には方向性の相違が生じる可能性もあり得るのではないかと考えている。これは南北問題とは別の観点において、意見の相違が顕在化する可能性があるということであり、そうした事態が生じた場合には、さらなる工夫が必要になるものと感じている。いずれにしても、今後の動向を注視してまいりたい。

【落合委員】
・今後グローバルサウス等への拡大を進めていくとのお話があったが、こうした国々においては、例えばソブリン AI 等の問題に関心を持たれている場合も多いのではないかと考える。この点について、議論において検討されている事項があれば、ご教示いただきたい。
・EU においては先月デジタルオムニバスパッケージが公表され、EU AI Act についても修正提案がなされているものと承知している。公表から間もないため、現時点では大きな変化は見られない

かもしれないが、こうした動向を踏まえ、国際的な議論において感じられていることがあれば、ご教示いただきたい。

【寺村特別交渉官】
・ソブリン AI については、日本としても検討すべき概念であると認識しているが、実際 G7 内においても、表立ってこれについて議論はされていない。おそらく総論としては賛成であるものの、各論においては多種多様な意見があるという状況だと思われ、現時点においてソブリン AI に関する議論が活発に行われているかというと、必ずしもそのような状況にはない。
・EU AI Act の動向についても、我々として注視しているところである。本件に関する個人的な所感を申し上げると、例えば中国において DeepSeek が登場した中で、欧州が同様の AI を開発しようとした場合、現行の EU AI Act が障壁となる可能性があると考えられる。こうした認識に至ったことが、現在の見直しの議論につながっている要因の一つではないかと推察している。ただし、この点については詳細な分析を行っている段階ではないため、今後様々な機会を通じて情報収集を行い、動向を把握してまいりたいと考えている。

【落合委員】
日本も、そうした意味においては、適切なバランスを保ちながら取り組んできているように感じる。ただし、こうした国際的な議論の中で得られる知見は多々あるものと思料するので、今後、日本としてもそうした知見を適切に取り入れていくことが望ましいと考える。

＜議事 4　AI事業者ガイドラインの更新に向けた論点に関して＞
【中川委員】
・AI エージェント、より正確にはエージェンティック AI と呼ばれるものについてであるが、誤動作が発生した際の損害をどのように取り扱うかという問題は、経済産業省が所管する民事責任の問題に該当するかもしれないが、極めて重要な論点であると考える。誤った予約や購入が行われ、損害が発生するといった事態は、今後頻繁に生じることが予想される。その対応策としては、私自身も現在研究を進めているところであるが、一つには保険の活用が考えられる。AI エージェントに付保する保険の導入や、より抜本的な方法としては AI エージェントを法人化するという選択肢もあり得るが、法人化した場合においても当該法人が保険に加入することとなるため、保険の在り方については避けて通れない問題として議論いただきたいと考える。誤動作の原因については、おそらく指示の仕方に問題があったか、あるいは AI エージェント側の能力が不足していたかのいずれかであると考えられる。その原因究明は非常に困難であるが、これは開発事業者の問題に直結するため、開発事業者に対して原因が把握しやすい形での開発を促していくことが極めて重要ではないかと考え、申し上げる次第である。
・エージェンティック AI は複数の AI エージェントにより構成されるとのご説明があったが、そうであれば、AI エージェント間でどのようにやり取りを行うかというプロトコルやコミュニケーションの方法に

ついても留意する必要が生じる。すなわち、人間が AI エージェントを使用し、相手方の人間も AI エージェントを使用する場合、AI エージェント同士がやり取りを行うという状況が想定される。その際に最も重要となるのは、文献調査等によれば、トラスト、すなわち信頼の構築であり、いかなる構造によりトラストを形成するかを明確にする必要がある。これは事業者のみの問題ではなく、政府においてトラスト構築の在り方に関するガイドラインを提示することが一つの方策であると考える。トラストという用語について、ご検討いただければ幸いである。特に、フィジカル AI との関係においては、対人 AI エージェントにおいてトラストが重要となる。例えば、介護施設において、入居者が一人になった際に対応する AI については、トラストがなければ適切に機能しないことは明白である。こうした観点から、トラストという用語を適切に取り込んでいただければ幸いである。

・リスクの問題について申し上げる。リスクベースアプローチは確かに困難を伴うが、逆に言えば予期しない事態も発生し得るため、インシデント発生時にいかなる対応策を講じるか、また、いかなる対応組織を構築するかについて、そろそろ検討を開始した方がよいのではないかと考える。そうすることで、実際にはアジャイルな対応も取りやすくなると考える。リスクへの対応のみならず、インシデント発生時の対策という観点を含めていく必要があると常々考えていたため、一言申し上げた次第である。

【林委員】

・AI エージェントに関して、これは AI の自律性に関する議論であると理解したが、従来の人間が指示し AI が応答するというモデルから、AI が詳細な提案を行い人間が承認するという AI エージェントモデルへ移行する場合においても、いずれのモデルにおいても人間が最終判断権者であり、目的そのものを AI 自身が設定するような、いわば目的自律的な AI はまだ登場していないという理解でよいか確認したい。また、近い将来において、そうした目的自律的な AI が登場する可能性があるかという点についてもお聞かせいただきたい。

・AI エージェントのリスクについて、ChatGPT のような従来の質疑応答型の AI であれば、AI はあくまで情報を提供する道具に過ぎず、人間が自ら理由を検討し行動を決定しているという前提を維持しやすい。これに対して、AI エージェントのように AI が提案を行い人間が承認するというモデルにおいては、承認が事実上形骸化し、行為の理由や選択肢の絞り込み等もすべて AI が実質的に支配することとなり、AI がある種の権威として機能するようになるのではないかと考える。これが必ずしも問題であると主張するものではないが、つまり人間は、自ら熟慮して選択する主体から、AI が設計した目的関数に従ってレールの上を進む主体へと変容してしまう可能性がある。そうなると、法秩序はこれまで、人間は自ら行動理由を形成し、それに基づいて自律的に行動し得る存在であるという前提のもとで、法的責任やリスクを構成してきたと思われるが、その前提に揺らぎが生じるように感じている。この点については、最近別の研究会においても議論したことがあるが、事務局としての認識をお伺いしたい。

【藤本課長補佐】

・ご指摘のとおり、目的に応じた自律度という整理があり得ると考える。ただし、当面は AI エージェントについて、目的は人間より与えられるものとし、手段に対する自律度を有するものとして整理した上で、今回の修文を検討してまいりたい。一方、今後導入が拡大していくと思われるマルチエージェントの環境においては、AI 同士がやり取りを行う中で、個々の AI の目的自体を他の AI が修正していくといった事態も想定される。そうした状況が生じた場合には、改めて目的に対する自律度にも配慮した修文を検討してまいりたいと考えている。したがって、現時点における AI エージェントは目的に関する自律性を有さず、目的は人間が設定し、手段に関する自律性を有するものとして検討を継続する。
・人間の判断についてであるが、ご指摘のとおり、将来的には人間が全く判断を行わず AI に委ねてしまうケースもあり得ると考える。しかしながら、自らの財産や生命に影響を及ぼすような重要な判断については、人間による判断を仕組みとして組み込んでいくことが重要であると考えている。もちろん、人間による判断の機会を過度に設けてしまうと、AI を活用する意義自体が問われることとなるため、そうしたバランスについても、AI 事業者ガイドラインにおいて留意事項として記載できればと考えている。

【瀧澤委員】
・エージェンティック AI について、定義に関する意見を述べさせていただきたい。エージェンティック AI と従来の AI エージェントとの相違点という観点において、「人間の介在を最小限に抑えながら」という表現を加えてはどうかと考える。最近、AWS を含め各社がエージェンティック AI のサービスを提供しているところであるが、必ずしもマルチエージェントが主軸となっていない場合もある。具体的には、MCP サーバー等の技術により、エージェントが複数の MCP サーバーから様々な情報を取得し、高度に自律的な判断を行った上で、実際にアクションを実行するという形態がある。複数回の問い合わせや、場合によっては他のエージェントを活用することもあり得る。これらにおいては、人間の介在が相当程度減少しており、この点が定義として重要ではないかと考えた。
・これに関連して、リスクについても言及したい。先生方のご意見とも共通する部分があると思われるが、人間の意思介在が減少しているという点において、当初人間が意図した指示に対して、結果は類似しているものの、その過程において全く異なる処理が行われていたという事態が生じ得る。我々はこれをアライメントリスクと呼んでいるが、こうしたリスクが発生する可能性があると考えている。補足すると、これはフィジカル AI にも関連するものと考える。単純な例で恐縮であるが、例えば二足歩行ロボットに対して「前に 3 歩進んでください」という指示を与えた場合に、その判断の過程において両手が動くことがあり得る。手が動くこと自体については、人間は問題ないと判断すると思われるが、これはロボットが自律的に判断し、バランスを取るために両手を動かしているものと考えられる。ただし、状況によっては手を動かさないでほしい場合もあり得るため、こうした観点からも、人間の介在が減少しているという点を定義に含めるべきではないかと考えた。

【山田委員】

・AI エージェントに関する 9 頁目の記載について、現時点では初版であると思われるが、対策よりもむしろリスクの分類について、各種文献や本日の意見等を踏まえ、ある程度時間をかけて精査し、整理していく方がよいのではないかと考える。例えば、誤動作や誤注文といった問題については、本日の議論でも言及があったが、これが左側の分類のどこに該当するのかが明確でない。こうした事象は悪意に基づくものではなく、いわゆるタスクにおけるハルシネーションが発生しているものであり、おそらく最も頻繁に生じる問題であると考える。こうした点も含め、また、トラストという言葉が出ていたが、AI エージェント同士が通信を行った際に、過度に情報を発信するエージェントを信頼してよいのかといった問題も、リスクとして顕在化するものと思われる。こうした皆様のご意見を総合して、この部分をしっかりと取りまとめていきたい。
・論点②のリスクベースアプローチの在り方について、私は資料 24 頁の②のアプローチがよいのではないかと考えている。ユースケースは 1 億件、1 兆件と無限に存在し得るが、その中から真にリスクの高いものを抽出することは、努力すれば可能ではないかと考える。具体的には、EU AI Act 等のハードローが整備されている国において、注視すべき又は要注意とされているものは参考になり得る。また、各種 LLM ツール等の利用規約には、禁止事項が相当程度記載されており、こうした情報源を活用することで、ハイリスクと考えられる領域をある程度定義できるのではないかと考える。ただし、そうしたハイリスク領域は全体の 1 パーセント程度であると思われ、残りの 99 パーセントをどのように分類するかは非常に難しい問題である。私見としては、自社のデータを使用し、自社内でのみ利用する場合、すなわちインプットもアウトプットも自社に限定される場合は、リスクは相対的に低いのではないかと考える。これも一つの分類の考え方であり、顧客から預かった個人情報や第三者の著作物等を外部から取り込む場合、また外部に出力する場合には様々な問題が生じ得るが、そうした運用を行わないのであればリスクは低いという整理も可能ではないかと考え、付け加えさせていただいた。

【佐久間委員】
・AI エージェントについて、責任や誤動作に関する部分は、既に委員の先生方からご指摘があった点と共通するため割愛するが、1 点、セキュリティの観点からコメントさせていただきたい。具体的には、エージェントのリスクに関して、既存の要素と重複する部分もあると思われるが、データガバナンスとの連動という論点を明記した方がよいのではないかと考える。AI エージェントは、従来人間が行っていたデータベースへのアクセスを自律的に実行するという側面があり、これに伴い新たな脆弱性が指摘されている。例えば、本年 6 月には「エコーリーク」と呼ばれる脆弱性が報告されており、これは人間が不正なメールのリンクをクリックしなくとも、AI が自動的に不正なプロンプト等を取得してしまうというものである。したがって、AI がどの範囲のデータベースにアクセス可能かという権限管理の観点が重要となる。この点については、既に「認証認可機能」や「権限乗っ取り」といった資料上の記載のうちでも想定されているものと思われるが、データベースへのアクセス管理という視点を技術的観点から明記しておいた方がよいのではないかと考える。これは本質的には RAG 等にも共通する部分があるが、AI エージェントにおいては特に複雑化するため、セキュリティの観点から

言及させていただいた。
・AI リスクの評価手法、すなわちリスクベースアプローチについて、事務局において整理いただいた内容には、大枠として賛同している。その上で、提示されているフレームワークの留意点と、追加で言及すべき点についてコメントさせていただきたい。まず留意点としては、既に議論に出ているとおり、リスクの性質によって定量的な評価が容易なものと困難なものがあることに留意する必要がある。例えば、AI の誤動作のように自社の売上に直接影響するものは、今回提示されている枠組みに落とし込みやすいが、外部ステークホルダーの人権侵害等については、その枠組みに適合しない場合がある。海外の投資家コミュニティ等においては、人権侵害リスクがサステナビリティの文脈で注目され始めており、今後は市場メカニズムとして、レピュテーションリスクや企業価値への影響という形で企業へのフィードバックがなされていく可能性もある。ただし、少なくとも現状においては、短期的な売上等に影響しない指標として軽視されうることから、こうした論点が欠落しないよう言及していく必要があると考える。その上で、追加で言及すべき点として、便益との衡量の必要性を挙げたい。リスクベースアプローチは基本的にリスク管理の在り方に関する判断であり、この枠組み自体に問題はないが、実際の AI 活用に係る意思決定においては、AI リスクの評価は ROI との比較衡量により行われるものと考える。活用推進を見据えると、便益との比較をいかに行うかという視点も重要となる。この観点については、経営戦略の文脈において、例えば ROIC との比較をいかに行うかという議論を「AI 時代の経営意思決定とガバナンス」という文脈で我々としても行っているところである。
・AI 事業者ガイドラインの活用促進策について、既にご説明いただいているとおり、検索性の向上は技術的に重要であると考える。具体的には、プルダウンやトグルの導入・HTML 形式の適用・チャットボット化等が実現できれば効果的ではないかと考える。また、民間においては、先ほどソフトローとしての活用という話があったが、自社がこうしたガイドラインにどの程度適合しているかという成熟度を測る指標「AI ガバナンスナビ」についても、AI ガバナンス協会において開発を行っているところである。

【高木委員】
・論点①に記載されている AI エージェント・エージェンティック AI・フィジカル AI について、今まで以上に、AI 開発者・AI 提供者・AI 利用者という 3 つの主体間の関係を分かりやすく整理することが重要になるのではないかと考える。現在、AI 事業者ガイドラインの定義においては、AI のモデル学習を行う者が AI 開発者であり、AI システムを開発し提供を行う者が AI 提供者とされていると承知している。しかしながら、AI エージェントやフィジカル AI に関しては、モデルの開発よりもシステム全体の設計の方が重要となるケースが多くなるのではないかと考えており、そうなると AI 提供者の位置付けが非常に重要になるものと思われる。この観点から、9 頁及び 19 頁に記載されている主体ごとの対策について申し上げたい。例えば、9 頁の AI エージェントに関する主体ごとの対策においては、プロンプトインジェクション対策のためのガードレール技術が AI 開発者の欄に記載されているが、これは AI 提供者としても重要な対策ではないかと考える。また、LLM・フレームワーク・サービ

スの最新化についても、AI 提供者に求められる事項ではないかと思われる。さらに、19 頁のフィジカル AI については、緊急停止機能の実装が AI 開発者の役割として記載されているが、これはむしろ AI 提供者が中心となって実施すべき事項ではないかと考える。このように、AI 提供者の位置付けが非常に重要になると考えているため、改めてご検討いただければ幸いである。
・論点③の検索性について、読みたい箇所を検索しやすくできるとよいと記載されており、それ自体はそのとおりであると考える。これはおそらくリファレンスとしての使用を想定した記載であると理解しているが、逆に言えば、自らに関係する箇所しか読まないということを意味しているとも捉えられる。そう考えると、先ほど 1 点目で申し上げた主体の整理との関係が非常に重要になるのではないかと思われる。主体との関係について、AI との関わり方という観点から、説明をより丁寧に加えていく方がよいのではないかと考える。例えば、最近の AI エージェントにおいては、フルスクラッチで開発するよりも、フレームワークを活用する、あるいはサービスを利用して AI を構築するケースが増えてきていると思われるが、その場合にガイドラインのどの部分を参照すべきかといった具体的な事例を、主体の整理との関係において明確にしていくことで、非常に使いやすいものになるのではないかと考えた。

【千葉委員】
14 頁のフィジカル AI のサービス事例について、様々な例が取り上げられているが、ここに記載されている自動運転から、記載のないものまで、フィジカル AI のサービスは本当に多様な領域にわたっており、今後もさらに増加していくものと考える。その中で、例えば自動運転のような極めてクリティカルなものから、リスクの比較的小さいものまで、様々なフィジカル AI が存在する。これらに対して、可能な限り共通的な観点からリスクの洗い出しや対策の検討を行うことは、非常に望ましいことであると考える。しかしながら、対象範囲が非常に広範であることから、リスクベースでの議論と実効性のあるガイドラインの策定においては、注意が必要ではないかと感じた。

【小塚委員】
AI エージェントに関する 9 頁目の記載についてであるが、事故が発生した際の民事責任については、別の会議体において検討することで差し支えないと考える。しかしながら、契約等を通じた民事上のリスク配分については、本会議体において主体的に記載すべきであると考える。具体的には、対策における上から 2 行目の記載において、ガードレール設計の横に、AI 提供者と AI 利用者にまたがる形で「ヒューマンインザループによる人間承認」と記載されているが、AI 提供者と AI 利用者では、その意味するところが全く異なるものと思われる。AI 提供者の側においては、人間による承認が可能となるような利用契約等の設計を行うべきであることを明記した方がよいのではないかと考える。

以上