資料１－３

事務局説明

「前回の議論のまとめ」

第1回WGでの議論（１）

１　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。

- リスク評価は、International AI Safety Report 2026（第１回資料２ P4）のAIリスクの分類の枠組みを念頭に置き、まずミスユース（悪用）とマルファンクション（誤作動）の両面に対応すべき
- ミスユースは主に利用者側、マルファンクションは開発段階で対応すべきもの、システミックリスクは損害発生時に問題となるもの、という分類が出発点となる
- エージェント化により、応答途中の想定外の挙動や、高い汎用性・能力の悪用が生じている。AIを一括りにせず、能力に応じて評価すべき。計算量に加え、能力による判断基準も必要
- 実際に観測されているリスクと、将来的に考えられるリスクとでは対策が異なることに留意。今なにが起こっているかというよりも、システム全体の検証や制御の観点で議論するのが良い。
- 深刻なリスクについて共通理解がない。リスクの議論と規律対象を分けて議論し、モデル能力や想定被害を整理すべき。
- 想定外のリスクも考慮し、共通の評価軸で優先順位を付けるべき。個々のリスクにルールを積み重ねるのは、実際のリスクに見合わない
- AI法の守備領域とそれ以外を切り分け、個人情報関連の法令や各種業法で対応できる部分はそちらで対応すべき。「なんでもかんでもAI法ではない」
- AI法、各業法、各政策の対応関係を整理するとともに、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
- 連続性を持つ問題群をどのように切り分けてリソースを投入するかが重要

※一部は複数人の発言を統合して記載

第1回WGでの議論（２）

２　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。

- サイバーとCBRNEについてはミスユース（悪用）とマルファンクション（誤作動）の両面で対応が必要。性的ディープフェイクや青少年保護も優先度が高い問題
- フロンティアAIは、悪用や予期せぬ挙動により非常に速く甚大な損害を生じ得る。ソフトローでよいかは再考の余地があり、海外事業者へのサンクションの裏付けも課題
- AIエージェントのデータアクセスや権限付与のリスクは様々な事業者が対応すべきもの。ただし、既存の個人情報に関する規律等でも対応し得る
- 既存法で、可能な限り対応する現行AI法のスタンス自体は維持すべきだが、16条は機能している部分とそうでない部分があり、業法が及ばないモデルの開発・実験段階等に規律がない。悪用は既存法令の執行強化で対応できる部分が多いが、制御喪失は今後さらに増える
- 16条は権利利益の侵害発生が前提で、制裁的でもない。「侵害のおそれ」の段階で報告徴求できるかが課題
- 下流の利用企業は基盤モデルの内部に入り込めない。サプライチェーン全体での責任分担と、下流への情報提供が必要
- 開発者が不明なオープンウェイトモデルへの対応も論点

※一部は複数人の発言を統合して記載

第1回WGでの議論（３）

３　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。

- EU AI法や米国州法の枠組みを参照し、AIの能力に応じたリスク管理、情報の収集・報告、透明性を確保すべき
- 一定の情報開示はハードローで行う段階にある。具体的な基準はガイドラインに委ねることとし、他方で営業秘密を保護するなど事業者の過度な負担を避けるべき
- 報告・勧告・命令の段階的アプローチを基本としつつ、国民の安全に関わる場合は緊急命令・直罰も検討すべき。また、「おそれ」段階での報告徴収や、報告の用途・共有先を検討すべき
- 義務付けの対象となるAIの範囲と段階（開発／実装／利用／被害）を切り分けて考えるべき。また、相応の義務を課すとすれば、対象範囲は自ずと限定される
- 対象者・求める行為の明確化と予見可能性の確保、既存制度との重複回避、海外事業者への実効性確保（国内事業者だけに負担が偏らないこと）、ガイドラインの乱立防止、報告のインセンティブ設計、情報共有枠組みの簡素化等が重要
- 利用段階でのミスユースについて、利用者側にも一定の役割を果たしてもらう必要がある
- 取締役会や倫理委員会等のコーポレートガバナンスも活用していく必要。
- 安全性評価は勿論のこと、国民保護の観点での研究開発、AISIの体制・能力強化が必要。また、また、国とAISIとの情報共有も重要。
- 国家間の協定等も活用しながら、情報収集の在り方を模索すべきではないか。
- 協議会の設置などによる官民連携の枠組みも模索すべき。

※一部は複数人の発言を統合して記載

総論の方向性（案）①

【委員からの御意見を踏まえた方向性（案）】

対応すべきリスクの範囲について

- 第一回の議論において、個別領域の法令とAIに関する法制度で対応すべき事象を切り分けるべき旨の意見が複数あったところ、AIに関する法制度で対応すべき領域は、

  ○ **特にAIに関する法制度で対応すべき必要性が高いリスク**

  の範囲を具体的に検討していく必要性がある。

- また、第一回の議論において、AIを一括りにせず、能力に応じて対応を検討すべき旨の意見や、法的な義務付けの対象範囲を限定すべき旨の意見があったところ、

  ○ **特にハードローで対応すべき必要性が高いAIの種別**

  を性能要件等から、範囲を具体的に検討していく必要性がある。

	特定のリスク	それ以外のリスク
一定以上の性能を持つAI	AIに関するハードローによる規律の必要性が高い	既存の法令やガイドライン等で対応
それ以外のAI	既存の法令やガイドライン等で対応	既存の法令やガイドライン等で対応

総論の方向性（案）②

【委員からの御意見を踏まえた議論の方向性（案）】

検討すべき制度的対応の対象について

- 第一回の議論において、AIの利用者規制・事業者規制を区別して検討すべき旨の意見が複数あったところ、以下のような整理に基づき、**事業者を対象とした制度的対応を優先的に検討**していくこととしつつ、利用者側においても、どのような対応・措置が可能かを包括的に検討していく。
- 事業者規制を行うにあたっても、研究・開発・サービス提供等の段階のうち、どの部分をどのように規律していくべきかを検討。

事業者規制
AIをサービスとして提供する者

- 既存法令（業法等）で手当がされておらず、対応が必要。
- 一定の義務をハードローで課すことを検討すべき。



利用者規制
AIを利用する者

- 犯罪行為の処罰や、個別の業法等における適正なAIの利用の推進は、既存の法令などでも対応ができる部分が多くある。
- 一方で、ハードローによる対応は馴染まない部分も多く、ソフトローも活用しながら、より一層利用者側のガバナンスも推進していくことも検討すべき。

総論の方向性（案）③

【委員からの御意見を踏まえた方向性（案）】

ハードローで規定すべき内容について

- 第一回の議論において、諸外国の法令も参考にしつつ、AIの能力に応じたリスク管理、情報の収集・報告を透明性を確保すべき旨の意見があったところ、ハードローの規律内容としては、事業者の義務、報告、勧告、命令、罰則等の規定等を法律の内容としつつ、具体的な内容を検討していく。

	特定のリスク	それ以外のリスク
一定以上の性能を持つAI	AIに関するハードローによる規律の必要性が高い	既存の法令やガイドライン等で対応
それ以外のAI	既存の法令やガイドライン等で対応	既存の法令やガイドライン等で対応

規定内容（素案）

- 要件に該当するAI事業者が特定のリスクに係るAIの安全性を確保するための措置を講じる義務を規定
- 国が当該リスクに係る情報を的確に収集するための、事業者から国への報告について規定
- AI事業者に対し、上述の義務を履行させるため、必要な措置を講じるよう勧告できる権限を規定
- 必要に応じて、国がより強制力のある命令（違反に対しては罰則で履行を担保）を行える権限を規定

（参考）高性能AIの規律に係る主な項目についての欧米法制比較(第１回WG資料より抜粋)

事項	米国(CA/NY/IL)	EU AI Act
適用対象	• Frontier Model=foundation modelでFLOPS数10^26以上 • Large Frontier Model=前年度売上5億ドル超 • ※LFMのみframework公表等の追加義務	• GPAI Model with Systemic Risk • ①High-impact capabilityを有するor ②個別指定 (51条1項) • FLOPS数10^25以上で①を推定（同条2項）
危険低減措置	• Framework実施義務 • リスク閾値、評価、レビュー、第三者評価等がframeworkに含まれる • ただし、具体的にどのようなmitigationを置くべきかは定めず	• モデル評価、レッドチームテスト等実施義務がかけられ(55条1項)、Commissionは自ら実施する評価に応じて具体的なmitigation実施を要求することも可(92条,93条)。
平時モニタリング(透明性確保)	• Framework10項目をweb上で公表し、年1回レビュー。壊滅的リスクの評価と結果等を公表する義務 • 内部利用による壊滅的リスク評価は3か月ごとに当局報告 • ILは第三者監査の実施と結果公表を要求	• Training contentの十分に詳細な要約の一般公表義務（53条1項d） • 当局リクエスト時はtechnical documentationの当局提出（53条1項a）
インシデント対応	• Critical Safety incident発生時に当局へ発見から15日 (NY/IL州は72時間)以内に報告 • 死亡又は重大な身体的危害の差し迫った危険がある場合は24時間以内に報告	• Serious incidentを不当な遅滞なく報告（55条1項c）。任意のCode of Practiceでは2~15日の期限が設定 • High-risk AI providerは原則15日以内に報告(73条)
実効性確保	• 違反・不提出時には最大100万ドルのcivil penalty(Attorney Generalによる請求。NY/ILは再違反時最大300万ドル)	• 違反・不提出時には最大世界売上3%相当or1500万ユーロの罰金(101条) • 最終的には市場撤退等を要求可（93条）

※主要な項目・規定の比較であり網羅的な表ではない。米国3州の定めは微妙に異なるがおおむね共通している部分を抽出。また、米国にはFTC法による強力な個別調査権がある。