Preferred Networks

資料2

AIの法制度について

Preferred Networks

2024年9月10日

想定されるリスク

- 生成AI一般に考えられるリスク
  - ハルシネーション、バイアス、権利侵害、倫理上の問題等

- AIの活用について当社で特に認識しているリスク
  - 当社は、自社AI半導体を用いた計算基盤をベースに、産業分野におけるAIの利活用（製造、プラント、バイオ・化学等）に注力
  - AIの種類・利用/活用方法・産業分野に応じた細やかなリスク認識が必須
    - 例：プラントの自律稼働でAIが想定通りに動作しないと、生命・身体・施設に重大な損害が発生する
    - 例：工場の稼働が止まると多大な経済的損失が発生する
    - 例：人命に関わる場面ではより高度な信頼性・説明可能性が問われる
  - 当社の取組については後ほどご紹介

リスクへの対応として、どのような制度の枠組が適切か

- **原則：既存の制度（事業者ガイドラインを主とするソフトロー）が適切**
  - 日本における立法事実が十分にあるか
  - 国際的な開発競争 ⇨ 一旦厳しめの規制を行いつつ社会受容を見て緩和するという方法論は、事業者の目線からは非常に厳しい（その頃には既に負けている）
  - AIの社会実装に対する評価（ AIで行うことが望ましい/許容される or 望ましくない）は定まったとは言えない
  - AI事業者ガイドラインで言及されている根拠については、引き続きあてはまる
    - 少子高齢化に伴う労働力の低下等の社会課題の解決手段として活用が期待
    - 法令の整備と技術発展・社会実装のタイムラグ
    - 細かい行為規制によるイノベーションの阻害
  - ガイドライン発出直後の立法は拙速 ⇨ 自主的取組の評価をまず行うべき
  - 日本企業のコンプライアンス意識の高さ（開発、ユーザともレピュテーション・事業リスクに配慮）
    ⇨ これらが機能しない場合に立法事実
  - 現状典型的に想定されている権利侵害（誹謗中傷、ディープフェイク、誤認逮捕等）についてはAIを用いた結果行為の段階で既存の法規制（名誉毀損、偽計業務妨害、刑訴法等）で捕捉可能
  - 安全性の観点からは、業法その他特別法や業界団体の安全基準等で対応

⇨ まずは整備されたガイドラインに基づいて、官民共同でベストプラクティスを作っていくべき段階（当社も対応）

リスクへの対応として、どのような制度の枠組が適切か（続）

- なお、モデルの規模に基づく法規制も適切ではないと考える
  - 社会実装についてはモデルサイズや学習量よりもデータの品質が重視され、実効性に疑問
  - モデル自体が社会に害悪を及ぼすのではなく、その使われ方によって社会に対するインパクトやリスクの程度が異なる
  - 日本における大規模なモデルの開発競争力を減殺し、日本はAI開発市場で後れを取る
  - 外国に所在する開発主体法人（ビッグテック）に対する実効性（執行可能性）に疑問

- **例外：国の安全保障、政治、経済、教育においては、国産LLMの使用を検討すべき（外国産モデルをベースにすることのリスク）**
  - 生成AIのアウトプットが、外国の思考、文化、言語等に強く影響、日本の立場、国益、伝統、文化、言語等を軽視したものになる可能性 ⇨ 安全保障、政治、経済、司法、教育の場面における国の情報発信や意思決定に対して（ステルスでも）悪影響、日本にとって適切な情報発信や意思決定ができなくなる可能性
  - 国の政治経済や安全保障におけるAIの利用を外国産モデルに依存すると、外交関係や計算資源の問題などで外国産モデルの利用ができなくなった場合、国の政治経済や安全保障に重大なリスク

リスク対応のための当社の取組（続）

- **技術的な取組の例**
  - 高品質なデータによる学習
  - ログチェックによる監視
  - フィードバックの収集による改善
  - 産業応用における（物理的）安全性の確保
    - 各産業分野における安全基準の遵守　例：プラント保安分野AI信頼性評価ガイドライン (https://www.fdma.go.jp/relocation/neuter/topics/fieldList4_16/pdf/r03/jisyuhoan_shiryo_03_02.pdf)
    - 取組先企業との協働　ユーザ企業が求める安全基準

- **法的な取組の例**
  - AIの種類・活用方法・分野に応じた機能制限
    - 例：汎用原子レベルシュミレーターのMatlantisでは対応元素に一定の自主制限
  - 適切な用途制限・禁止事項の設定
    - 例：PLaMoでは、信用・教育・雇用・法律・医療等々の分野に係る重要な決定など、個人・企業にとって法的又は重大な影響を与える可能性のある目的におけるアウトプットの使用を禁止
  - 規約に違反するユーザの調査