経済産業省 第 5 回 AI事業者ガイドライン検討会
議事要旨

令和 7 年 2 月 20 日(木)
10:00～12:00
オンライン会議

冒頭挨拶

- 第 5 回 AI事業者ガイドライン検討会では、委員の皆様からいただいたご意見を踏まえて、事務局として対応案をご提示する。さらにご意見を賜りたい。
- 令和元年に「AI・データの利用に関する契約ガイドライン 1.1 版」を公表したが、生成 AI の普及を始めとする近年の市場環境の変化を踏まえ、契約に関する当事者間の適切な利益及びリスクの分配、AI の利活用を促すことを目的として、「AI の利用・開発に関する契約チェックリスト」を 2 月 18 日に公表した。是非ご参照いただきたい。

◆ 議事(3) 全体討議

■ 2-1.AI によるリスクの洗い出し・分類

- リスクの分類について、政策的な側面が強く、具体的なアクションを記述しないと、事業者がイメージできないのではないか。例えば、選挙等に関する記述は政策的なものであり、事業者のリスクに相当するものとはいえず、削除が望ましいのではないか。
- 「雇用の喪失」は事業者には馴染まないリスクと考えられる。リスクは明確に整理すると良いのではないか。
- リスクの分類について、何がリスクであるか等のリスクの明確化が必要であり、「データや利益の集中」は政策リスクといえるのではないか。また「雇用の喪失」は、政府の課題と認識されるが、雇用を支援することは事業者の責任でもある。「雇用の喪失」だと事業者に考慮すべきリスクが伝わらない可能性があるため、リスクの名称は「労働者の失業リスク」等とし、そのリスク対策として、労働者支援プログラムとしてリスキリング等を含めると良いのではないか。
- リスクの分類について、「雇用の喪失」、「データや利益の集中」は、たしかに直接的に事業者のリスクではないが、重要なリスクではあるため削除ではなく、事業者のリスクとして認識してもらえるように文言を例えば、「労働者の失業リスク」、「データの集中」にするとよいのではないか。
- 「労働者の失業リスク」のリスク対策として、リスキリング、リテラシー向上を入れると良いのではないか。
- ポジティブリスクに関する記述がないので、ポジティブリスクも含めてどこまで対応すべきかの議論が必要なのではないか。
- 政府としては AI のリスクのメッセージペーパーとなるため、リスクに関する注意書き等は必要なのではないか。例えば、全てのリスクを記述はできないので重要なものを記述していると思うが、本当に重要なものなのか、リスクについて議論が行

われている中で、日本国内で必要とされるリスクをしっかり記述すべきである。また、「悪意」等のリスク分類に偏っているのではないか。

- リスクの分類において、「過度な依存」を、倫理・法に関するリスクとして分類していることに違和感がある。人間中心と言っている以上、人間の身体・精神に対する項目を設けて、その中に「過度な依存」等を盛り込めないか。
- 各モデルも提供形態によりリスクは異なることから、正確に検討する必要があるのではないか。
- バイアスに対する懸念の高まりや近時のプログラム用途での LLM 利用に即したリスクについて事例を追加してはどうか。

■ 2-2.AI の契約に関する留意事項

- 業務委託契約における AI の利用が増加傾向にあり、明示せずに AI を利用しているケースもある。発注者側から AI を利用していることを明示しても良いのではないか。
- 合理的な説明の実施はどのようなリスクがあるか等の共有のみではなく、リスクの説明に加え、対応の説明及びどのような対応を講じるかが必要ではないか。

■ 2-3.生成 AI に関する記載の追加

- ディープフェイク等に関して、国際的にデジタルレプリカの肖像権に関して法的な位置づけについて触れると良いのではないか。
- バイアスに関しては、セキュリティ（情報漏洩）に関する注記、出力に文化的なバイアスが入っている注記を記載することが良いのではないか。
- 外部サービス・データと連携する場合における個人情報流出の可能性を記述しても良いのではないか。加えて、OSS（オープンソースソフトウェア）の使い方も指摘したほうが良いのではないか。
- モデルによっては、出力結果にバイアスが含まれる場合があり、現時点では AI 利用者はどのようなモデルを利用すべきか判断ができず、提供者が判断している状況である。目的に応じてモデルの利用を補記することが良いのではないか。
- 入力するプロンプトのバイアス、RAG のデータ等のバイアスに関して考慮するべき旨は反映されると良いのではないか。

■ 2-4.AI ガバナンスに関する事例の充実

- ガイドラインの活用向上に向けては、家電の取扱い説明書のように数ページで基本ガイドのようなものが必要かもしれない。例えば、経営層・リスクマネジメント部・法務部・開発現場等から成る AI ガバナンスチームの確立、社内の AI の開発・利活用状況の把握、AI ポリシーの作成、（AI の開発や利活用に関する）行動基準の作成、開発や利活用に関する手順を整えていくことなどを盛り込んだガイドである。「AI ポリシー」のひな型の提示や、フローチャートによる開発者・提供者・利用者の判別などができるとよい。また、説明会の実施や企業（国内における海外事業者も含む）へのヒアリングを通して、啓発を進めるのがよいのではないか。

■ 2-5.AI ガバナンスの動向等の反映

- 米国での最近の議論として、AI の大きなインシデントがある場合、AI の監督・監視に責任を負うことが挙げられており、この点を注記することが良いのではない

か。また、AI ガバナンスに関する社内組織の設置に加えて、取締役会における AI ガバナンスの監視を促す注記があるのが良いのではないか。
- 外国法のコンプライアンスリスクについて本編で述べても良いのではないか。
- 技術をオープンにすることによるリスクと透明性とのトレードオフのバランスが難しいことに関しては、サイエンス等の論文を参照し、コラム等により補足するのが良いのではないか。

■ **2-6.特定単語の整理・見直し**
- 開発よりもどのように利用するかに焦点が移っている。責任の観点からみると、提供者の役割が大きくなっている。そもそも開発者、提供者、利用者の分け方で良いかという疑問もある。また、ユーザビリティや提供者にフォーカスすべきという問題意識があり、リビングドキュメントとして再構成をする必要があるのではないか。
- 「学習」及び「開発」は多義的であり、見直しが必要ではないか。
- 「学習」は多義的な意味があるため、誤解がないように補足することが良いのではないか。AI 事業者ガイドライン（1.01 版）の別添 P5-6 の図 2 及び図 3 は、AI の学習過程を示す図だが、これはディープラーニングであり、In-context Learning 等は継続して学習しているわけではない。「学習」に関しては、長期的に情報が保持されるかどうかで、深層学習と文脈内学習を分けて考えるべきではないか。強化学習アルゴリズムのバイアスも考える必要があり、GPT3 時の RLHF（人間のフィードバックによる強化学習）が公平な手法かと問われたら回答が難しい。
- 現在の AI 事業者ガイドラインは、「訓練」と「学習」を使い分けているわけではないため使い分けるのが良いのではないか。また、バイアスの対象は少数の集団には限らないため記述を更新するのが良いのではないか。
- 「学習」に関しては、広義ではモデルの挙動を変えるもの全体を指す。事前学習だけではなく、In-context learning も入力に応じて挙動が変わり、例えば KV キャッシュを付け加えれば「学習」と言え、今後はむしろそういった学習が増えていくと思われる。また、2025 年は強化学習を使うケースが多くなり、現行もしくは今後のモデルを使用することで誰にでも強化学習が可能となると考えている。
- 「開発」、「学習」、「データの種類」等に関して、読者の誤解を招かないように整理が必要なのではないか。

■ **2-7.その他**

➢ **データ関連**
- 安全性に関して、学習データが違法かどうかは、膨大なデータもあり人の判断が難しく、技術的な確認も厳しい。学習データに違法なデータ等を含めないことに留意するという表現ではなく、「可能な限り努める」旨の記述をすると良いのではないか。
- 学習データ等について海賊版の利用に関して、海賊版と知りながら学習する場合は規範的侵害主体として責任を取られる可能性が高まる点、より丁寧に記述するべきでないか。
- 個人情報保護として、子供のデータや公開されているデータの利用について議論されているため触れると良いのではないか。

➢ **ボリュームや UI**
- 現在のガイドラインはボリュームが多く、ガイドラインの詳細を説明する AI チャット

作成やガバナンスに関するチェックリスト等がある良いのではないか。

- ガイドラインの複雑化は避け、更新方針を定めたほうが良いのではないか。
- ISO 等の他のガイドラインとの関係と AI 事業者ガイドラインの位置づけの明確化をすべきではないか。
- 消費者庁の AI 利活用ハンドブックがわかりやすいものの、具体的なものがわからないため、その内容を AI 事業者ガイドラインにおいて補足すると良いのではないか。
- ボリュームに関しては委員の方々のご指摘のとおりであり、リビングドキュメントの更新方針を検討するのが良いのではないか。

以上