

第2回分科会における議論（情報セキュリティ等関係）の概要

第1 全般

個人情報保護・情報セキュリティ及びAIシステムの整備・調達・運用・セキュリティを中心に、短期的な想定ユースケースのみならず、中長期的な想定基本機能も念頭においた議論がされた。

第2 AIモデルの選択等（裁判所業務関係）

1 AIモデルの選択に当たって考慮すべき事項について議論がされ、まず裁判所において独自のAIモデルを開発することの必要性や実現可能性については、事前学習から独自で行うことのみならず、独自の追加学習をさせることも含め、裁判所において要件定義等を行う必要があり、これらに要する費用や人材の確保等の観点から、現実的ではないということで異論はなかった。

2 次に、裁判所における利用目的や用途に照らして、当該AIモデルが必要な性能を満たしているかどうか重要であり、想定ユースケースとして挙げられている争点整理の補助といったものであれば、大手外国企業が開発したAIモデルは必要な性能を満たしていると評価することに異論はなかった。これに加えて、国内企業が開発したAIモデルはこれらよりもやや性能が劣るのではないかと考えられるが、プロンプトの設定が出力の精度に大きな影響を与える面がある上、性能は日々進化しており、将来の裁判所における導入時点においては十分な性能を有していることも考えられるとの意見があった。

また、司法権が国家主権の一部であることや裁判官が外部から不当な影響を受けずに法と良心に従って職権を行使するという裁判官の職権行使の独立との関係を考慮する必要があることに異論はなかった。

これらの点は、究極的には司法権の行使に対する国民の信頼の確保に帰着するから、裁判所におけるAIの利活用についても国民の信頼が得られるか否かが重要である、近時のソブリンAIの考え方（一般的には、AIの開発・学習・運用に必要となるデータ、計算資源・インフラ、モデル、人材などを可及的に自国内で調達・管理し、国外への依存を抑制しようとする政策・方針のことをいう。）も考慮する必要があるのではないかと、いった意見があった。

これらの点からは、裁判所において利用するAIモデルが「外国製」であるか「国産」であるかが問題となりうるが、そもそも「外国製」といっても、外国企業が開発したことを意味するのか、機械学習が外国で行われ

たことを意味するのかなど、様々な観点があり、結局のところ、利用目的や用途との関係で当該ＡＩモデルを利用することが司法権の行使に対する国民の信頼を揺るがせるものではないといえるかが重要であるとの点で異論はなかった。

そして、裁判所が「外国製」のＡＩモデルを採用したとしても、最終的に裁判官が出力結果を確認することを前提とするのであれば、「外国製」ＡＩモデルについて指摘されることがあるバイアスの存在といった観点を考慮しても大きな問題はないのではないかと意見が出されたほか、特に、争点整理の補助のような想定ユースケースにおける利用を前提とした場合には、我が国の司法権の行使に外国企業が不当な影響を及ぼすおそれがあるとは考えにくいということで異論はなかった。

また、裁判所におけるＡＩモデルの選択について、国民の信頼を確保するという観点から、透明性確保のための仕組みやガバナンスが必要となるのではないかと、ＡＩモデルの評価については政府における検討状況も考慮すべきではないかと意見があった。

- 3 最後に、ベンダーロックイン（特定のベンダーの技術やサービスに依存し、他社への切り替えが困難になる状態をいう。）や開発企業の撤退等によるサービス提供停止といったリスクを考慮して、ＡＩモデルを選択する必要があることで異論はなかった。なお、複数の種類のＡＩモデルを利用（乗り換え）した結果、出力が異なる可能性があるものの、最終的に裁判官が関与することが確保できれば、特に問題はないと理解できるのではないかと意見があった。

第3 情報セキュリティ・ＡＩセキュリティ（裁判所業務関係）

1 情報セキュリティ・ＡＩセキュリティ全般

裁判所業務において利用されるＡＩシステムにおいて求められる情報セキュリティ・ＡＩセキュリティは、一般的な行政事務において求められる情報セキュリティ・ＡＩセキュリティと大きく異なるところはないと考えられることについて異論はなかった。

もともと、将来的に裁判所におけるＡＩ利活用が国民の権利義務に直接影響を与えるようになってきた場合には別途考慮が必要になってくるとの意見があった。

2 ＡＩモデルの配置場所等

前記第2の2のとおり、国民の信頼の確保との関係にも留意すべきであり、セキュリティ確保に関する説明責任の観点からはオンプレミス（システムの稼働やインフラの構築に必要となるサーバ、ネットワーク機器及び

ソフトウェア等を自社で保有し運用するシステムの利用形態をいう。)が望ましい側面はあるものの、調達・構築・維持管理のコストや導入期間の長期化等の課題があることから、裁判所が管理する日本国内サーバのクラウド上に配置することが総合的にみて最も妥当ではないかとの意見があり、異論はなかった。

3 AIセキュリティ・AIモデルの公表

裁判所におけるAI利活用でのAIセキュリティについては、特にプロンプトインジェクション攻撃、プロンプト窃盗攻撃等への対応を検討すべきであるが、一般的に求められる対策を講じることが重要であり、裁判所独自の対策を講じることまでは必要ないのではないかということで異論はなかった。

これに関連して、AIモデルの公表は、一般には透明性と予測可能性が高まることや今後のAIモデルの調達における公平性の確保にもつながるメリットがある一方、攻撃者における脆弱性の発見等のおそれ等のデメリットがある旨が指摘されたが、現在広く利用されている汎用の生成AIに対する攻撃のリスクについては、近時はそのような事例は見当たらず、大きな問題とはならないのではないかとの意見があった。

なお、裁判所において使用するプロンプトについては、個別の事件におけるプロンプトの使用方法を裁判所の外部に公開することにより、担当裁判官の判断過程や方向性が使用するプロンプトから推察される可能性が高まり、裁判官の職権行使への影響が問題となり得ることから、非公開とすることで良いのではないかとの意見があった。

4 生成AIシステムの調達

裁判所の生成AIシステムの調達を検討する際においても、現在改定作業中の政府ガイドラインに準拠した検討が必要であり、かつ、それで足りるということで異論はなかった。

5 生成AIシステムの監査

一般に生成AIシステムに対する監査としては、RAGの導入状況、AIモデルの設計の適正性、AI利用者による適切な利用（適切なプロンプトの設定を含む。）、ファクトチェック、緊急時対応等を総合的に取り入れた監査が必要とされ、このようなAIの監査基準は整備されつつあるとの指摘があった。

もっとも、例えば、個々の事件における裁判官のプロンプトの使用方法について監査することは、当該裁判官の判断過程の評価を伴うことになり、裁判官の職権行使の独立との関係で問題が生じうることから、裁判所においては一般的な監査基準によることは相当ではないとの意見があり、

現状では、裁判所の生成ＡＩシステムについて、ＡＩの学習・推論・生成におけるすべての過程の適正性に係る監査証跡を保存してそれに監査（特にリスト型）を実施することで客観的に示すことは困難であることについて異論はなかった。そして、裁判官の判断の適正性を担保するという観点からＡＩガバナンスの検討を深める必要があることで意見が一致した。

なお、プロンプトに関しては、上記のと通りの監査の視点のほか、裁判官が使用しているプロンプトを裁判所内部では相互に共有することによって、ＡＩの出力の精度を上げていくことに活用できるのではないかという意見がある一方で、裁判官の職権行使の独立との関係での検討を深める必要があり、例えばプロンプト全体を収集するものの、個別の事件での判断過程がわからなくなる程度まで抽象化して共有するといった手法も含めて検討を深める必要があるといった意見があった。

6 ＡＩガバナンス

前記第２の２のＡＩモデルの選択や前記５の生成ＡＩシステムの監査に関しては、ＡＩガバナンスの在り方についても議論が及び、裁判所が利用するＡＩモデルがその利用目的や用途に沿った機能を有していることや利活用の際に生じるリスクについて、定期的（少なくとも年１回以上）に裁判所において検証評価し、その結果を裁判所内部において共有するとともに、ＡＩの利活用の方法の改善やＡＩモデルの更新の場面等で活用することが必要であり、このような検証評価の結果は、その概要だけでも法律実務家との間で共有し、透明性を確保することが必要ではないか、裁判所に特化したＡＩの情報セキュリティマネジメントを構築し、第三者による定期的な評価を行うことで説明責任を果たすべきではないか、裁判所全体において、裁判官へのリテラシー研修を実施することにより、裁判官一人一人のＡＩに対する知識や経験を維持向上させるとともに、個別の事件においてＡＩ利活用の適否が問題となったとしても、当該事件の担当裁判官の心証形成の問題として上訴等の訴訟法上の不服申立手続において解決することで、当該事件の判断の質を担保するという全体の枠組みを構築することで、個々の裁判官がＡＩの利活用に躊躇することがないようにしていくことが望ましいのではないか、といった意見が出された。

第４ 弁護士・司法書士業務関係

一般論として、セキュリティあるいはサイバーセキュリティに留意する必要があるとの意見があったものの、固有の課題についての指摘はなかった。