AI ネットワーク社会推進会議
AI ガバナンス検討会
第６回　議事概要

１．日時
平成 31 年３月５日（火）10：00～12：00

２．場所
中央合同庁舎２号館　第３特別会議室

３．出席者
（１）構成員
平野座長、江間構成員、大屋構成員、河島構成員、木村構成員、木谷構成員、久世構成員（代理：日本 IBM 株式会社 立花 東京基礎研究所 AI 担当シニアマネージャー）、小塚構成員、三部構成員、城山構成員、杉原構成員、中川構成員、長田構成員、原構成員（代理：富士通株式会社 中条 デジタルサービス部門エグゼクティブディレクター）

（２）総務省
山崎大臣官房総括審議官、竹内サイバーセキュリティ統括官、泉国際戦略局審議官、
井上情報通信政策研究所長、香月情報通信政策研究所調査研究部長、
市川情報通信政策研究所調査研究部主任研究官、
高木情報通信政策研究所調査研究部主任研究官

（３）オブザーバー
実積中央大教授（OECD デジタル経済政策委員会（CDEP）副議長）、
内閣府、消費者庁、個人情報保護委員会、文部科学省、情報通信研究機構、理化学研究所、
産業技術総合研究所

４．議事概要
（１）事務局からの説明
机上資料１に基づき、構成員からの意見について報告があった。

（２）有識者からの発表
資料１に基づき、以下の発表があった。
○「AI ベースシステムの事業化における課題」（日本電気株式会社　セキュリティ研究所　谷幹也　所長）

（３）事務局からの説明
資料２及び机上資料２に基づき事務局より報告があった。

（4）意見交換

＜NEC セキュリティ研究所 谷所長の発表について主な意見＞

【江間構成員】

・　モデルからデータが復元できる可能性があるという話だが、説明可能性や透明性、バイアスの話も含めると、本当に正しいモデルを作るために学習させたのかということを見なければいけないということと、しかしながら中のモデルは守らなければならないということの、トレードオフの問題を含む気がする。ブラックボックスに対する透明性、説明可能性の話にも関わってくると思うが、どうお考えか。

【NEC セキュリティ研究所 谷所長】

・　トレードオフの問題があるのはおっしゃる通りだと思うが、ここで言いたかったことは少し異なる。機械学習をしたモデルはパラメータセットが並んでいるようなデータセットになっているだけなので、学習に使用したデータを復元できるとは思っていない人が多かったが、実は復元できる可能性があるということを示唆したかった。

【江間構成員】

・　目的に沿ったものになっているかということが重要だと思う。ものによっては偏っていても良いわけだが、社会システムに使うのであれば、偏りがあるといけないものもあるかもしれない。

【NEC セキュリティ研究所 谷所長】

・　説明責任上、モデルがどのようなデータから出てきているかということを統計的に処理して見せることができれば一番良いと思うが、今はそこまで技術が進歩していない。
・　そういうことを判断して決めている会社かどうかということで認定基準を作り、その AI については問題ないということを言わなければいけないのではないか。今はそのレベルでしか解消の方法がないと思っている。

【中川構成員】

・　幾つかの AI が企業に属していると、企業秘密という表に出てこない状況下で競合や協調が起こることがある。その場合、複数の AI が織りなすアウトプットの場みたいなものを観察・観測しているようなタイプの AI が一番求められると思う。

【NEC セキュリティ研究所 谷所長】

・　あくまでも外部で調整するというのは、それぞれの企業の AI が出してくる結果を外部観測した上で判断できるレベルでしかないと思う。外部観測から得たものに対しての示唆あるいは指示をするような AI は必要で、重要インフラの場合は特に重要になってくると思う。

【城山構成員】

・　データのところだけでロバストにするのは限界があるので、社会システムである程度まで担保するとなると、応用分野ごとにある種の社会的システムを作らなければいけないことになり、それは必ずしも当該企業の倫理だけではなく、分野全体の社会システムの在り方までも含めて「ここまで」と決めざるを得ないと思う。そういう意味で品質保証はどう行うのかということをお伺いしたい。

【NEC セキュリティ研究所 谷所長】

・　「これを超えてはいけない」という仕様はきちんと守りつつ、その中で AI が自由にできるというのが、安全な AI の使い方であり品質保証ができる。AI が今まで思ってもみなかった連携をすることでより効率的な動きをするということまで求めるのであれば、その時の品質保証は分からない。
・　データをどれぐらい処理して作ったものだから安全だという基準はできるかもしれないし、AI を作る時にどういうインプットを使ったかということの評価の仕方はあると思う。
・　認定制度の適用を企業に委ねるだけで良いとは思っていない。悪意のある人間を押さえる仕組みが今はないので、それをどうするかというのが課題だと思っている。
一つは、仕様としてどのように外に枠がはめられるのかという基準で品質保証をするというやり方と、もう一つは、作り方に関する基準を設けるというくらいしか品質保証はできないと思う。

＜事務局の説明について主な意見＞

【三部構成員】

・　保険の利用に関して、EU においては、信頼されるための AI の概念と倫理ガイドラインの中でもアカウンタビリティの原則の中に入っている。
・　自動運転については既に保険利用の検討が始められており、実務的には保険金が支払われる事由の定め方をめぐって難しい議論が起こっている。安全性の確保がかなり高く求められる分野については、保険の必要性が高くなると思う。
・　逆に保険の利用を強制してしまうと、保険会社の側として利用しづらい面があるだろう。その意味で“措置の例”としていることは、選択の余地を認めていることだと理解できるので賛成できる。
・　第三者機関の設置についても、例えば EU のガイドラインの中で ethical board の設置ということが言われている。

【中川構成員】

・　セキュリティ侵害時の書道措置の点だが、AI が使われる場面を考えたときに、リアルタイムで使われるタイプと、オフライン的に使われるタイプとでは異なるのではないか。リアルタイムで使われるシステムであれば、レジリエンスという概念を入れなければならない。オフラインの場合は、むしろ精密に原因を究明し、時間を掛けて被害の大元を正すということをする。初動措置においてこの違いは意識しておいたほうが良いと思う。
・　プライバシー侵害時に講ずるべき措置の例として「拡散ルートの特定、保存先消去の依頼」

とあるが、インターネットを経由して出ると拡散ルートの特定はほぼ不可能に近い状況になる。誰が持っているかということもほとんどわからなくなるので、個人情報保護法にある提供元基準を意識されたほうが良いと思う。

【小塚構成員】

- セキュリティ侵害時の措置の例として保険が書かれているが、事が起こってからでは保険は利用できない。セキュリティでは被害者に対する補償賠償をすることがおそらく求められる措置で、それを円滑あるいは十分に進めるために保険の利用が推奨されるということだろう。
- 某保険会社が自動運転に対応した保険商品を出したという記事が出たが、それは実際には第三者に対する保険ではなく、初動措置などをするための保険である。その意味で、保険の利用というのはもう少し他の部分にも係るものかもしれず、書き方は要注意である。
- 同じことをプライバシーの原則に書かれるとさらに面倒になる。そもそも、プライバシー侵害のときに第三者への補償賠償ということを書き切って良いのかという問題もある。
- 「プロファイリングを行う場合には対象者のプライバシー等に慎重に配慮する」とあるが、プロファイリングがプライバシーの問題かどうかということ自体が大きな論点なので、ここにプライバシーと書き切って良いのか。「対象者に生じ得る不利益に配慮する」などという方が中立的かもしれない。

【平野座長】

- 安全の原則のところには保険の話は書いていないが、セキュリティの原則のところで書くならこちらにも必要ではないか。

【小塚構成員】

- むしろ安全の原則の方が、三部構成員がおっしゃった EU などで推奨している保険の話になるのかもしれない。ここにも「保険の利用などが救済のために有益である」ということを書いていただくと良いのではないか。

【NEC セキュリティ研究所 谷所長】

- セキュリティ対策の対象が「AI のセキュリティ」と言われてもわからない。今日講演したのは AI そのもののセキュリティのことであるが、ここに書かれているのは AI サービスに対することだと思うので、「AI のセキュリティ」と書くと少し語弊があるように感じる。
- 工場などのオペレーションでは、セキュリティだからと言って、システムの停止やネットワークの遮断がすぐにはできないということが今一番の課題になってきている。すぐに止められないものに対してどのように対応するかを併記しておく必要があるのではないか。

【中川構成員】

- 「データ提供者についても、データ形式の標準に準拠する」とあるが、形式だけではなく、特徴を含めてお互いに理解していないと、AI 相互間や他システムとの連携はできないと思

う。必要なデータを取っていなかったということが多くあるが、そのときに形式だけでなく特徴の共通化、共有化ということも技術的に必要になってくるので、お書きいただいた方が良いと思う。

【実積オブザーバー】

・　AI ネットワーク化により惹起・増幅される課題への留意のところで、無数の AI がネットワーク化によってお互いが相手の挙動を見つつ調整していった結果価格をフィックスしてしまうというケース（AI 談合）に対処することを考えているのであれば、少数の AI の影響力が強くなるというよりも、最終的に独占をしているのと同じような状況が AI のコラボレーションによって生じてしまう、ということを踏まえた書き方をした方が良いと思う。

【城山構成員】

・　「セキュリティの原則」の各論点のところで、一方では「セキュリティ侵害時」という書き方をしていて、他方、セキュリティ対策時のためのサービス提供等のところでは「インシデント情報」と書いているが、この「セキュリティ侵害時」と「インシデント」をどう整理しているのか。ここは少し明確に表現をしたほうが良いと思う。
・　第三者機関の設置という議論とも絡むが、すべてのセキュリティ侵害時が対象になるというよりかは、重大なことが起こった時にこういうオプションがあるということだと思うので、むしろそこに至らないようなものが、インシデント情報の共有をまずプロバイダなり何なりが行って、彼らはそれをサービス提供者にも提供しなければならない、という構造になっていると思う。言葉の定義と同時に、全体の構造を明確にしていただいた方が良い。
・　Adversarial Example のところで「学習に利用されたデータが不均衡な場合に攻撃が存在する可能性」となっているが、論理的には、悪意に基づく攻撃があった場合に学習に利用されたデータが不均衡になり、結果として当該学習モデルが誤った結果をもたらすリスク、という順序になると思う。

【久世構成員（代理：日本 IBM㈱東京基礎研究所　立花　AI 担当シニアマネージャー）】

・　データ形式の標準や共有というところは、ビジネスモデルに関わり特徴量で共有するのは難しいケースもあると思うので、今の書き方で良いと思う。
・　消費者的利用者がデータ形式について情報を踏まえた上で収集、保存を行うということについては、ガイドラインに書くほどのことでもない気がする。収集したデータを何に使うかはそれを集めた人とそれを利用する人の間の問題である。

以上