個人情報保護法のいわゆる３年ごと見直しの
検討の充実に向けた事務局ヒアリング
議事概要

１　日　時：令和６年 12 月３日（火）13:00～
２　場　所：個人情報保護委員会
３　出席者：
（１）ヒアリング対象者：
一般社団法人 AI ガバナンス協会
国立研究開発法人情報通信研究機構（NICT）
一般社団法人日本ディープラーニング協会
プライバシーテック協会
（２）個人情報保護委員会事務局：
佐脇事務局長、西中事務局次長、小川審議官、吉屋参事官、香月参事官

４　議事の概要
（１）ヒアリング対象者からの説明
①AI ガバナンス協会から、資料１に基づき主に以下の点について説明があった。
・協会内で、各企業のAIガバナンスへの取組状況を可視化する取組「AIガバナンスナビ」の開発・運用を進めている。初回の調査では、個人情報の取扱いについて、適正にできているという自己診断をしている企業が多かった。
・ただし具体的な運用を調べてみると、業務で生成AIサービスを利用する際の個人情報の利用等について、企業によって規制への解釈やどこまで保守的に運用するかにばらつきがあることがわかる。実態としては保守的な方針に倒して進めている企業が多く、活用は道半ばと思われる。
・今回の見直しに際しては、AIをはじめとする統計的な個人データ処理がこれまで以上に常態化していくということを前提に制度を考えていく必要がある。その中でイノベーションと権利利益の保護の両立が重要なのは言うまでもないが、原理を唱えるだけではそれらの価値の調整を図ることは難しいので、具体的な活用シーンを一つずつ検証していく必要がある。
・その上で、個々の活用シーンと向き合いながら制度検討を行う上では、①リスクベースアプローチ、②技術中立性、③マルチステークホルダーアプ

ローチという３点の考え方を原則として検討を進めるべき。

- （参考４[1]）１①・②・⑤・⑥・⑧　本人同意を必要とする規律自体に一定の見直しの余地がある。同意よりも利用目的や態様に応じて個人データの取扱いの正当性を裏づける基準を導入するという方向にシフトしていくほうが現実的ではないか。その理由は大きく二つあり、一つは、そもそも権利利益の保護という観点で見たときに、本人同意の実効性の限界が、特に情報量が増えてきている中で出てきている。加えて、事業者側でも同意取得の手続がかなり重く、利用形態を少し変えるだけで毎回同意を取るのは現実的ではない部分があり、そこは合理化する余地がある。実体的ルールの議論は先述の三つの原則に則り、様々なステークホルダーを巻き込んで進めていく必要がある。
- （参考４）２　技術的にも保護が可能な部分は本人同意を個別に免除する余地がある。AIモデルの学習を目的として、要配慮個人情報等を含むデータを用いる場合や複数組織の個人データを突合して活用する場合は、安全管理措置の部分を議論した上で、一定の権利利益保護が果たされていれば同意のない利用を認める余地がある。データ管理のフェーズで言えばPETs（プライバシー強化技術）等を活用したデータの加工、モデル開発以降のフェーズで言えば学習モデルから個人データが引き出されないようテスティングやガードレールツールの導入など、技術的な手段で権利利益の保護が果たされれば、個別の同意を要件とせず、個人データの利用を認めることも考えられる。
- （参考４）１④　既に取得されたユーザーのデータセットをAIモデルの学習に利用する場合、事前の通知の内容によって目的外利用になり得るケースがある。そもそもの本人の権利利益に関わる利用目的の通知をしっかりするのは大前提として、モデル学習に用いる際に毎回同意取得を義務付けることは技術中立でなく、開発の遅延にもつながりうる。ここは事業者からの声が大きい。
- （参考４）３　生成AIサービスを利用する際のプロンプトに個人情報が含まれるといった論点について、事業者間でも解釈が異なっている部分があり、この部分がどういった要件の下で認められるのか、議論し明確化していく必要がある。
- （参考４）４　AIのサプライチェーンの複雑化を背景として、大規模言語モデルやそれをベースとするソフトウェア構築を請け負うベンダーとそ

[1] （参考１～４）について、第 310 回個人情報保護委員会　資料１―１「『個人情報保護法のいわゆる３年ごと見直しの検討の充実に向けた視点』に関するヒアリングの概要について」別添１の参考１～４を参照。

の利用企業の関係で、データベースの構築や処理は第三者に依存するケースが増えており、各事業者にどの程度コントロール可能性があるかを踏まえ、責任関係の整理が必要。いきなり制度化するよりまず実務的なスタンダード形成が重要。

・（参考４）６　インプットする個人データの内容に即して規制のレベルを切るというやり方では、必ずしも現実的な権利利益の保護に寄与しない場合もあり、AIモデルの学習等に用いる場合と併せて、どういった規律が望ましいか検討が必要。

②情報通信研究機構から、資料２に基づき主に以下の点について説明があった。

・大量の日本語のデータで学習した国産 LLM は必要。皆さんが使っている海外製生成 AI の学習データは英語中心で、日本語データは僅かしか入っていない。その結果、日本の主張、文化、アイデンティティーが海外の LLM によってかき消される可能性がある。海外製の LLM を使うと、子供が好きなおもちゃについて我々が普段聞かない海外製のものを答えるなど明らかに英語データで学習した影響が見られる。こういうものが使われるようになると、我々の文化はどうなるか心配になる。また、いわゆるデジタル小作人に落ちぶれて国富が流出する状況も想定される。

・NICTでもLLMを作っているが、フェイクニュースをつくらせると簡単にきわどいものが出る。LLMがマネタイズできるかは議論されているところだが、仮にマネタイズできなかったとしても、今後、LLMはいわゆる認知戦での主要兵器として色々な組織、主体に使われることになるだろう。

・一番深刻な問題は、質はともかく、生成情報の量では、生成AIは人間社会全体を凌駕しつつあること。実際に２か月で１億人と対話した生成AIがあるが、人間には生物学的に不可能。誰かが本気でフェイクニュースを生成AIで大量につくりSNS等にポストした場合、人力での対策は不可能。近い将来、日本社会は生成AIで守るしかない時代がくる。フェイクニュース、誹謗中傷、マルウェア、現在想定できない何かが生成AI発でやってくる可能性があり、生成AIによる反論でもってフェイクニュースを無効化するなどクリエイティブな方法で社会を守る必要がある。

・生成AIが正義にのっとった行動を常にするとは限らず、複数の正義を志向する生成AIで互いのネガチェックをしながら、あるいは議論をしながら社会を守っていく施策が必要になる。

・LLM、生成AIでは学習データは極めて重要。我々NICTでは、過去15年間にわたって大量のウェブデータを収集。日本語に関して言えば生成AIの学

習でよく使われるCommon Crawlという一般に公開されているデータの５倍程度のデータが含まれる。日本では生成AIの開発が遅れているが、日本の数少ない勝ち筋になると思っており、現在、生成AIを開発する企業や公的機関等に提供する試みを開始している。

・データの提供は個人情報保護法を踏まえ学術研究を目的に含む共同研究の枠内に限定。人材獲得競争が非常に激しく、共同研究に当たることができる人材の確保が困難。さらには共同研究をすると手の内をさらすことになりデメリットがある。共同研究では、NICTの側でもカウンターパートとして人的含むリソース等が必要で、多数の主体を相手に共同研究を行い、データ提供をすることは困難。共同研究を前提とせずに、要配慮個人情報を含むウェブデータを民間に提供できれば社会を守る意味でも有益。

・ウェブデータは一般公開されているものしか集めていないが、ここから個人情報を完全に特定し削除するのは技術的にほぼ不可能。

・プライバシー保護と生成AIの実用化、サービス化の両立を目指すのであれば、学習データから個人情報を削除することが望ましいのか疑問。まず、データから個人に関する情報を全て抜くと学習する知識が歪み、出来の悪い生成AIしかできない可能性がある。より現実的な方法は、生成AIのサービスを提供する側で、個人名等が含まれる入出力をさせないような手段をサービス段階で提供すること。例えば、サービスを開始した後に学習済みのデータからの削除漏れが見つかった等の理由で学習済みのデータから個人情報を削除するとなると、削除後に学習をやり直す必要が生じ、場合によっては数十億円かかる。そういったリスクがあると、恐ろしくて生成AIの開発投資もできない。サービスの段階でフィルターをかけるなら要請に応じ氏名等のブラックリストをつくればいいのでよりローコストで確実。

・加えて、学習データから個人情報を一律に抜くと、生成AIは氏名等が認識できなくなり、さらには何が誹謗中傷か認識できなくなる。例えばネットに誹謗中傷が大量に出て、この対策をAIでしなければならないとなった時に、個人情報を抜いた学習データで学習した生成AIは全く無力になる。つまり、社会を守るという点でも非合理。

・共同研究なしでウェブデータを生成AIの開発者に渡すことが認められると、データ提供に際しては、提供先の民間企業と提供元の公的機関は契約を結ぶこととして、個人情報を含めて不適切な利用を行わないこと、情報漏えい対策を講ずること等の義務を提供先に負わせることを検討する必要がある。社会的・安全保障上のリスクもあることからデータをオープンに誰でもダウンロードできる形にすることは考えていない。

・これまで、研究目的でウェブデータを収集していたため、そのデータを企業側がビジネス段階で活用することが難しい。生成AIを使ったビジネス開始後も学習済みのデータを手元に置きトラブルシューティングあるいは著作権侵害のチェック等をしたいというニーズがあるもののそう言った理由で活用困難。
・ビジネス開始後も学習済みのデータを手元に置けないと海外勢との競争で非常に不利。結果として国産の生成AIで社会を守る、文化を守ることが画餅に帰す。この辺に何らかの配慮をいただきたい。

③ディープラーニング協会から、資料３に基づき主に以下の点について説明があった。
・（参考４）２　要配慮個人情報取得の規律について、データベース等を構成するもののみに限定するか、少なくとも生成AIの学習のような一般的・汎用的な分析結果の獲得と利用を目的とするものは除外すべき。例えば公開データセットやクローリングで収集したデータを用いてデータセットを作成し、このデータセットを学習したモデルを開発し、その学習済みモデルとデータを公開したいというニーズがある。
・個人情報保護委員会のOpenAIに対する注意喚起において、例えば収集する情報に要配慮個人情報が含まれないよう必要な取組を行うことや、収集をできる限り即時に、できる限り減少させるための措置を講ずることを求められているのが現状。想定ケースについて、個人の権利利益にどれほど影響があるか。そもそもウェブサイト上に存在する要配慮個人情報の大半は、本人や報道機関が掲載しており大半は取得時の本人同意が不要。実際にCommon Crawlのデータの病名や犯罪名が含まれるテキストデータ8,000件超を目視確認したが、大半は本人による公開や、報道内容。実質的に見て本人同意が必要そうなものは僅か３件程度。こういったケースを類型化すると、①親族等、自分以外の身近な人の病歴について記載されているもの、②誹謗中傷で社会的身分や人種等を用いているもの、③少年犯罪に関して実名を記載しているもの、を発見した。注意喚起に従ってできる限り減少させても、こういったものをピンポイントで抜き出すのは困難。
・全部削除するには広くフィルタリングが必要。病名や犯罪名などのデータベースを突合し、該当するものを全部削除することはありえなくはないが、過剰なフィルタリングによりデータの絶対量が落ちてしまう。また、病気や犯罪に関して極端に弱いモデルが出てしまう問題がある。また、誹謗中傷などで使われる病名や人種名が正式名称でなく別称も含んでおり、

それを全て排除するのは過剰なフィルタリングをしてもなお難しい。

・そもそもクローリングで取得するデータは膨大で、ファイル形式も通常のものと異なり、一般の人の目につく可能性は基本的にない。Common Crawlは2007年頃からクローリングデータ公開を広く行っており、これを加工して公開しても個人の権利利益への新たな影響は基本的にない。

・生成AI開発では学習データの規模が重要で、クローリングデータを使う必要性が高い。例えばCommon Crawlのデータセットをダウンロードし日本語モデルをつくるときには、いろいろな言語のデータが入っているため日本語部分を抜き出す必要がある。またホームページから抜き出すので、会社概要や、文章ではない用語も入っており、そういったものを使えるようなデータに加工し、場合によっては加工したデータを公開したいというニーズがある。

・生成AIの開発投資が限られている日本企業では、個社で海外のビッグテックのような開発を行うことは難しく、技術のオープン化もある程度は重要。その場合、開発したモデルを公開すれば良いという話ではなく、再現性の確保が必要で、開発したモデルだけでなくデータセットも併せて公開したいというニーズが強い。

・そもそも、クローリング技術自体は検索エンジンなども含め広く用いられている。Common Crawlのデータを利用する場合、一括ダウンロードすることになるので、注意喚起で求められている、収集する情報に要配慮個人情報が含まれないように必要な取組を行うことは難しい。

・海外の企業がOpenAIに対する注意喚起を踏まえ、これを本当に守っているのか、実効性が疑問。他方で、この注意喚起は日本企業への影響は大きい。収集する際もできる限り減少させることとされており、どこまでやるか各企業はかなり真剣に検討している。特にデータを公開するとなると慎重にならざるを得ず、保守的な対応がなされている。

・（参考４）２・３　AI学習は、委託や利用目的の通知・公表等の規律において、統計処理と同様に扱うところを明確にしていただきたい。想定ケースの一例としては、顔画像、ビデオデータ等の個人データを含むデータを入力し解析を行うAIを組み込んだSaaSを提供する企業において、当該SaaSのユーザー企業から入力された個人データを含むデータを、AIの学習に用いたいというニーズがある。学習に用いる場合、①ユーザー企業１社のデータを学習するパターン、②複数のユーザー企業のデータを合わせて学習するパターン、③複数のユーザー企業のデータを、個人データを突合したうえで学習するパターンがある。現状は、①はできるが、②ができるかどうかは必ずしも明確ではない、③は明確にできない。②について、

SaaSでは、学習用や解析対象のデータとして実際に複数のサービスユーザーから受領した個人情報を一つのデータセットとして学習や情報解析に供する必要性が高く、AIの学習を統計処理と同様と考えれば認めない理由もないため、既存の規律でいう委託の枠組みが使えることを明確化していただきたい。③についても、プライバシーテック等の活用により一定の場合に行えるとする立法の余地はあるのではないか。

・企業が持つ有用なデータが技術発展により取得時に想定していなかった分野のAI学習に用いる場合もある。推論を用いるなら利用目的に記載する必要があるところは当然だが、学習において必ず利用目的の通知・公表が必要かは疑問。統計処理と同様、個人の権利利益に必ずしも大きな影響を与えない。

・（参考４）３・４　生成AIの利用の局面について、いわゆるクラウド例外かそれに準じた例外か、また委託、越境移転についての適用範囲や要件の明確化が必要。また、サービス利用者のみに責任を負わせるべきでない。サービス提供に当たり複数事業者が開発するものについて、これらサービスの利用において現状のクラウド例外や委託の規律だけでは対応が難しいところがある。特に一つのサービス内で複数の生成AIのモデルを利用できるサービスも増えており、それを利用するときに、ユーザー企業で利用する生成AI全てについて利用規約を確認し、個人データの取扱いについて何に影響するか調査する必要があるのが現状だが、そもそもクラウド例外の該当性や、委託、越境移転の規律の判断が難しい。同じサービスについて多数の利用者が同じ審査をやっており無駄がある。サービス利用者において個人データを入力する可能性があるSaaSを利用するときに、どういった法的枠組みで、どういった要件でできるのか明確にしていただきたい。かつ利用者側で要件該当性の判断のための情報の取得を容易にすることがAIの利用促進につながる。

・法務は理解しているが現場が理解できていないことが多く、実際に提供元基準や容易照合性を現場の一従業員は理解しておらず、意図せず法令違反が生じているケースがありうる。

・ヨーロッパ等でデータ利活用が進んでいる現状を踏まえ、ドメインごとの特別法等の検討がより進むとありがたい。

・AI技術について、問題が起こった論点について指針を示していただくのはありがたいが、問題が未だ起こっていないものについても検討が必要なものは多い。現状、AIに関する個人情報保護法の論点の明確化のニーズは大きいが、ガイドラインやQ&Aの中にAIに係る記述があまり入っておらずニーズと比例する形になっていない。

④プライバシーテック協会から、資料４に基づき主に以下の点について説明があった。

・（参考４）２・３　同意なしで提供した上で突合と分析をし、統計利用するのであれば、権利利益を侵害しないのではないか。そのとき、PETs のうち特に秘密計算はデータを秘匿しながら処理でき、秘密計算を使うと権利利益を保護しながらの活用ができるため、積極的に検討いただきたい。

・PETsがどういったところで活用できるかというと、例えば医療機関が電子カルテデータを、ヘルスケア事業者が運動履歴情報を、それぞれ秘密計算を使って秘匿化しながら突合したデータを作り、それを使って統計的な分析をして、統計情報を出力するようなものである。なお、学習モデルも統計情報の一種と考えている。

・次世代医療基盤法においてもヘルスケアデータは対象外。こういったユースケースには社会的意義があるが現状は対応できない。疾病や運動量の情報は、匿名化して属性を曖昧化すると疾患分析の精度が落ちるため困る。

・幾つか方式があり、TEEやMPCは標準が存在。TEEはハードウェアレベルで安全性を確保した秘密計算で、処理中でもメモリ内データが見えず秘匿した状態にできる。GPUにも対応し、この数年で海外での導入が進んでいる。

・統計的利用や一般的・汎用的な分析を行うことに限定される場合は同意不要。ただ、統計的利用、一般的・汎用的な分析を担保することが重要。その担保方法として秘密計算が合理的で効果的。こういった技術を積極的に推奨することが望ましい。

・この検討を急ぐ必要がある。人手でこれを担保するのは難しく、委託先の管理が難しく漏えいした事故が数多くある。また、海外のガイドラインはここ１、２年で非常に多く出ておりこの技術の導入が進んでいる。この技術が日本で強いのはあくまで研究開発レベルで、事業レベルでは非常に遅れている。アメリカでは、この１年で国防領域にこの技術を入れるという動きもあり、クラウドでも日本企業は残念ながら入っていない。これを進めていくことは経済安全保障の観点でも重要。

（２）各ヒアリング対象者と事務局との主な質疑応答は以下のとおり。

AI において許容されるアウトプットについて　※「参考資料１－１（参考４）２」関連

（事務局）

○　本日提起された問題の一つは、データ処理の結果のアウトプットが、ある種一般的で特定の個人をターゲットにしたアウトプットではないという意味で、統計などと通底する類似性があり、したがってそこにインプットされるものは、名簿等の場合とは発想を変えても良いかということだと理解している。個人の権利利益ないし個人のデータをどう個人がウォッチするかという点に着目して議論を組み立てる論者の場合には、通常のデータを使われるシーンがそうであるように、自分のデータは基本的には自分の予想の範囲で使われることが重要という議論になる。

今回、AIや統計のインプット部分について、個人の同意その他の規律の緩和の議論をしようとした場合、一つ考えられるのは、個人が同意その他で関与できるようにする範囲について、自分のデータ全体ではなく、「自分のデータを用いて自分に直接フィードバックされるデータ処理に関しては、引き続き自分の予想の範囲で使われるべきもの」と定義し直し制度を組み立てるかどうかという考え方。その限りにおいて、自分のデータ全般について自分の予想の範囲で使ってほしいという個人の希望を否定する政策判断になることだと理解している。要配慮個人情報について現行制度上は同意が前提になっている。その背景は、その処理によるアウトプットの多くは、確率論として差別性を誘引するきっかけになるという考えのもと、それに基づくデータ処理を用いた活動が、多かれ少なかれ差別的なことに使われ得ることを前提に、本人に対して事前に同意を求めていったということ。AIなどで要配慮個人情報があるがゆえに出てくるアウトプットは、別に要配慮個人情報のインプットをコントロールしたとしても、差別性のない状態が確保されるかというと、それが効果的なコントロールの仕方ではないというのはそのとおりかと思う。ただし、本人からすれば、引き続き自分のデータが使われているがゆえにどんなアウトプットか関心があるという点は、意見として残るだろう。

そうなると、結局幾つかのことをAIのアウトプットにおいて確認し担保することを含めた制度設計になる可能性がある。まず、データをインプットする局面では、統計その他の特定の個人から離れた一般的な知見をアウトプットするために使うという点について、その遵法性をどう担保するかという観点がある。次に、一般的な知見ということでアウトプットはされるものの、その内実は、アウトプットが世の中にさらされた瞬間に、特定の人が確実に差別されるような体系的な知見である可能性もあるので、そういったアウトプットが本当に出てこないことが保障されるかという、AIなり、統計処理のアウトプットそのものに関わる規律をどうする

かという観点がある。最後にそれをクリアし、許容できるアウトプットだとして、それを個人の名簿と突き合わせて本人に影響をもたらすような活用が想定されるから、それが許容されるかどうかという観点もある。これら三つは少なくとも挙げられると思うが、個人情報保護法は、最初のインプットの観点については、インプットされた個人データが一般的な知見を得るだけのために適正に利用されているかという点を含め、利用目的の通知・公表といった規律で一定のガバナンスを効かせられる。あと、最後の観点、すなわち、AIその他のアウトプットを使って、個人に影響を及ぼすような個人データの利用のシーンは個人情報の取扱いにほかならない。しかしながら、AIのアウトプットが、許容されるものであるかどうかという観点については、個人情報保護法から引き出される施策の範疇にはあまり入らないと思われるため、最終的に、どんなAIだったら良いかという議論は残ってしまう。

もしかすると、許される利用目的は何かをリスクベースでどう評価するかということにも近しいと思われるが、私どもはそこを決め切る事前の知識もなく、一体どうしておくのがいいのかよく分からない。その辺について、イメージなり、御示唆なりコメントいただけると幸い。

（プライバシーテック協会）

○　学習データを使って学習モデルをつくり、それが統計情報であれば同意なしでやれる。その学習モデルを使って個人の名簿と突き合わせて、個人に対して何かアクションをするなら同意が必要と理解。学習モデルが本当に統計情報かどうかについて、そこは技術の進展などにもよるが、モデルから学習データが出るおそれなど様々ある。そのため、この技術的判断を個人情報保護委員会がするのは難しい。大枠は個人情報保護委員会が示し、技術的な部分は専門家が必要。ある程度民間と連携し決めていくと良い。

（事務局）

○　質問を補足すると、こんな議論がある。統計、AIのようなデータ処理のために個人情報をインプットする場合、同意は要らないという議論もあるが、今の個人情報保護法で第三者提供の同意規律を緩和しているのは、基本的に公益目的のような、個人の利益を超える目的が掲げられる場合に限定し、その限りにおいて個人に我慢してもらっているのであるという考え方もある。「我慢している」と言った趣旨は、本来、いかなる場合にも、自分の情報をどう取り扱うかについては自分が納得すべきである、という立場の方々の受け止め方を表現してみた次第。一方は、本人による同意の機会を制約し得るのは公益目的がある場合に限られるという立場、

その反対には、一般的な知見やデータ処理、通常、統計と言われるものであれば、個人には関係ないと割り切ってはどうかという立場。仮に、両者の中庸を取ろうと思うと、「こういう一般的知見等であるから、アウトプットの問題はないです」と言ってあげる必要が出てくるかもしれないが、何をもって問題がないと言えるか、あまり知恵がない。

EUのAI Actのように、禁止されるAI等のカテゴリーを設定するという議論が日本では本格的に展開される状況にはないと理解しており、どういう対応があり得るのか、ご意見を伺いたい。

（情報通信研究機構）

○　これまでの議論は、既に一般公開されているウェブデータがメイン。そもそも「自分の情報」とは何を指すのか明確化して議論した方がいいのではないか。ネット上では本人に由来しなかったり、本人がオーナーシップを主張できる様なものではないもので、本人に関する情報が多数ある。個人情報保護法が本来想定していると見受けられる個人情報とウェブ上の個人に関する情報は性質や来歴が違う。そのため規制の在り方も、別に考える方が望ましいのではないか。

出力に関しても、今の生成AIでは差別的な発言をすることはあり得るが、それが個人に紐付いていなければこの場で議論の対象になるような話ではないとも考えられる。そもそも、学習段階は正直我々技術者でも中で何が起こっているのか分からず、精密に分析してなにか言えと言われても困る類いの話なので、そこに何か規律を課すのはあまり現実的ではない。一方で、出力の直前あるいは入力の直後で個人の名前や住所、電話番号を出さない、受け付けないといった処理は、より確実性をもって行える。これも100%完璧ではなくベストエフォートにはなるが。よって提案としては、最終的なサービス段階で自分の情報がどう使われるかに関して御理解いただけるようなサービス形態となるような規律なり何なりを課せばよいのではないか。

（事務局）

○　ウェブ上で拾ってきたデータを特定の目的のために個人情報データベースにする場合には、本人は当該データに関し、事業者に対して一定の権利を発動できるというのが個人情報保護法の建前。したがって、ウェブ上のデータがデータベースになる瞬間に、そのデータベースは、本人が意見を申し出る客体になる点を、どうすべきか悩ましいところ。その延長線上で、生成AIサービスについて、本人から要求があればオプトアウトに応じているところもあると思うが、学習データその他でインプットされるデータについて、それを気にする本人がいた場合、その人たちの権利をケアし

ようと思う場合の最も典型的な権利設計の仕方は、オプトアウトをジェネラルに認めるという、分かりやすい方法があるが、今の話だと、それは無理だという御主張か。

（日本ディープラーニング協会）

○　そこで言うオプトアウトとはどういうところか。

（事務局）

○　自分の名前を検索し該当したら学習済みのデータから全部削除するということ。

（日本ディープラーニング協会）

○　既に学習したものについて学習をし直すことはまず不可能。「学習データから削除して今後使いません」であれば、技術的には対応できる場合もなくはないように思われる。ただ、データセットとして公開していると、ほかで流通してしまう問題はある。

（情報通信研究機構）

○　生成AIでビジネスをやっている組織でも、学習済みのデータから申請のあった名前を抜いて再学習することはやっていないところもあるのではないか。コストが相当かかる。なおかつ、例えば、今申請を受け付けたとして次の学習が終了するタイミングはいつかという話がある。学習は恐らく数週間はかかるので、その間、データが使われたままのサービスが継続してしまう。十分なリソースがないところは、申請後、それがモデルに反映されるのは、学習済みのデータからの削除であれば下手をすると１年後とかになりうる。

　また、オープンでモデルを公開していて、すでに誰かがダウンロードしているケースでは要請の反映のしようがない。その辺を考えると、実際につくったモデルをサービス投入する際に、入出力に自分の名前があったら単純にブラックリストに載せてもらう方が合理的ではないか。ただ、その辺も気持ちの悪い問題はある。自分の宣伝をしたい人気商売の人と同姓同名の個人がいて、ブラックリストにその氏名を登録した後に人気商売の方の人が自分の情報が消えてしまったがどうしてくれるのかと言ってくるとか。そうした事態もある意味自分の情報をどう活用されるか把握しているかということだと思うので、簡単にやり直しが利く形が一番望ましい。個人情報データベースについては、要するに個人名が表形式になった瞬間に制約をきつくすべきなのはおっしゃるとおりだと思うが、生成AIを開発している人たちのほとんどは、別にそれをやっていない。禁止事項として、AI学習用のデータをもらってきたときに、そこから個人情報の表をつくるなとか、あるいは個人を特定して何かするための解析は

するなとか、そういったことは規律として書けそうだが、それはどうか。個人情報である・ないにかかわらず、AI学習を一般的に制約するのは難しそう。

（事務局）

○　おっしゃることは一つのアイデアだと思う。共同研究、あるいはその後、民間企業を含めて提供するという状況になった場合、ある程度相手を選ぶとか、契約で担保するとか、そういう規律を大事にしたいという話があった。制度論を考えたときに、そうして集めたデータを誰かに提供し、提供先がどう使うかという意味で、本日の議論に沿えば、そこはAIに食わせるのであれば、データは本人に、という議論は一旦リセットされて提供されることになる。ここで、提供する元、つまりデータを集めてきた人にとって、集めたものは何だと思えばよいか。個人データだと理解すればよいか。つまり個人データでもないと思うと、そこから先の提供は、提供先がちゃんとしていれば、何でもない行為だと制度的には評価することができる。一方で、提供先が何をしているかについて、一体どういうガバナンスの構造で社会全体として規律するかと言う点は大きな問題があり、たまたま個人データであったら、個人情報保護法の一般規律が一気通貫で適用されるが、AIやその類いに関連したデータトランジションに着目した、新しい常識みたいなものをスクラッチでつくり始めないといけないのかということがよく分からない。

一旦個人情報が観念的に含まれているので、我々は要配慮個人情報を含めて、個人情報に関する制度の現状の一般ルートとの差分で、どう設計するかという頭の体操を必要に迫られてやっているが、一体どんな世界が本来いいのか。その場合、データの提供元が提供先に対して責任を負うことについて、NICTはたまたまデータを集めておりかつ公的機関なので、そのように思っていることは理解するが、一般的にはどうするのが望ましいのか、どんな議論になると思うか。

（情報通信研究機構）

○　極めて難しい問題で、誰でも好きなことを発信できるというインターネットの今の状況では、情報の信憑性を含めて、いろいろな混乱が実際に生じている。まず指摘しておきたいのは、ウェブデータは、誰でも見られるデータ。なので、書いているほうはそれなりの覚悟を持って書いているだろうと思うが、ただ、その際に、プライバシーなどに思慮が至らないで書いている等のケースもあり、それが、今、問題になっていることだと思うが、そうした問題全体を一気に解決するというのは非常に難しいだろう。一方で、個人情報保護法をつくられたときに念頭にあったのは、そも

そも誰が誰から情報を集め、どういうデータにするか、きっちりトレースできる世界を考慮していたという印象をいつも受ける。インターネットになってしまった瞬間、そういったトレーサビリティーが一切なくなっているがゆえに、いろいろな問題が生じており、これを一気に解決するのは難しいのではないか。なので、個人とその属性を表にしたらアウトだとか、そういった類型を幾つか並べて、これはアウトということを潰していく以外に、個人的には思い浮かばない。

もう一点、AIの開発者のほとんどは、別に個人情報を出したくてAIをつくっているわけではないことにはご留意いただきたい。一方で、生成AIを日本企業がつくれないことにより、プレゼンで述べたように国全体で様々なデメリットが生じる。それはある意味公益に準じたようなものになるとも思われる、それらのことを考え合わせると、今までの個人情報の取扱いとは違う観点で整理が必要なのではないか。

（ディープラーニング協会）

○　先ほど想定していたクローリングの例でいうと、結局、散在している情報なので、取得ができれば、公開自体は現行法だとできるということになると思われる。公開されたとしてもデータの量が莫大なので、学習以外には使えず、個人情報を抽出するといった使い方はあまり想定されない。そういう使い方をする場合には、適正利用などで、例外的なケースについて網をかけていくことはあり得る。なお、公開した場合、転々流通して、流通先が悪用したらどうするかという問題はあり得るが著作権法第30条の４の議論は参考になりうるのではないか。同条では表現された思想または感情を享受する享受目的だと使えないが、非享受目的だと使えるため、著作物が含まれたデータセットを公開するときに、非享受目的に限定してくださいという利用条件を付けて公開することもある。個人情報保護法上許されない目的で使うことは不適正利用に該当しうるので、問題状況は似ているのではないか。転々流通は不適正利用に該当し得るので、例えば公開するときに一定の契約上の義務を負わせた上で公開をすることはあり得るのではないか。著作権法第30条の４と個人情報保護法の規律がある程度整合すると、使う側からは非常に使いやすい制度になる。

（プライバシーテック協会）

○　そういった規制を技術的に担保することも考えられる。海外の事例では、技術的に制限をかけている例もある。当協会からは提供後にデータを突合する提案をしたが、提供後に処理する内容や処理後にデータを削除することを技術的に担保可能。そのような規定を設けることも考えられる。

（事務局）

○　そこまでいくと、NICTがおっしゃるような、誰に提供するかという部分は、管理されたほうが心穏やかか。

（情報通信研究機構）

○　我々のデータで学習したAIでどこかの誰かが大量のフェイクニュースをつくって、SNSに投稿されても困るということはある。

（事務局）

○　それは本人が自分のデータをどう使いたいかという気持ちとあまり変わらないのではないか。

（情報通信研究機構）

○　国全体の安全保障に関わるような問題なので、個人の問題とは少し違いがあるように思う。

（事務局）

○　国全体の安全保障に関わるフェイクニュースに使われるかどうかということについて、それを気にしてコントロールするということか。それはあるかもしれない。

（情報通信研究機構）

○　あとは、あちこちで言っている話だが、最近、指示型の強盗、闇バイトがある。気がついたら、指示をしているのがAIということもあるかもしれない。そういう恐ろしい話が今後出てくると思うので、そういう悪いAIを作れないようにする。それは個人の情報が悪用される、されないとはちょっと差があって公や公共性が出てくるところではないか。

（事務局）

○　自分が思うようにデータを使ってほしいという思いと、ここでの議論を調和させると、その上で、日本を転覆させるために使ってほしくないと思う個人の意思は許されると思われる。そうすると、個人かどうかはともかく、どんなAIは許されるかという議論に結局コミットせざるを得ない。

（情報通信研究機構）

○　生成AIの学習で私の情報を使ってくれるなと思ったとして、私一人の情報を学習データから抜いてもらったとする。それは多分ほとんど影響が無い。よって、それが何人ぐらいになったら影響が出てくるのかという話も絡んでくる気がする。

（ディープラーニング協会）

○　おっしゃる問題は非常に重要ではあるが、個人情報保護法でカバーすべきことなのかという疑問はある。

（事務局）

○　もちろん、個人情報保護法ではないところだと思う。

（プライバシーテック協会）

○　透明性は必要なので、一定の説明責任を課すことはあってもいい。ただ、公益かどうかの判断はやはり難しいので、それを基準にするのは難しいと考える。

（AIガバナンス協会）

○　今の話も、例えば安全保障などの今議論されている問題そのものは、そもそもほかの公益に照らして駄目だというだけの話であり、技術中立性の観点でもAIを個別に想定した規律は不要ではないかと思う。先ほどの「個人のコントロール権をどこまで認めるか」という議論に戻ると、大きく三つ論点があるように思っている。

まず一点目は、個人が自分のデータの使われ方を完全にコントロールする、ということの実益は何かという点。そもそもどの程度実態としてのニーズがあるのかから議論が必要かと思う。例えばAIの学習に限らず、自分の名前が他者で噂話をされている程度のものの場合、権利利益の保護の対象に基本的にならないはずである。すると、「AIで学習されているかもしれない」という抽象的な懸念だけを理由に、個人の完全なコントロール権を認めることに実益はあるのかが問題。この点がクリアに説明されなければ、AIだけが技術中立でない形で規制される形になり得る。一般的な統計利用はある程度客観的に見ても権利利益の影響が小さいため、ゴールベースの考え方をとれば、本人同意の規律から除外し得る、というのが現状の見解。

二点目として、「よいAI」「安全性の高いAI」をどう定義するかは確かに難しい問題だが、一定の技術的な安全管理措置などを基準にすることは可能。例えば、AIソフトウェアから個人の名前と要配慮個人情報が関連付けられて出力されるような事態が生じた場合にはたしかに問題であり、学習データを制限する方法でこれを防ぐことはやや難しい。ただ、NICTのプレゼンでもあったように、インプット、アウトプットのレイヤーでこうした出力を防止するテスティングやガードレールといった技術は存在し、先ほど話に出たように、100%ではないものの、一定程度リスクを低減させることは可能。そうした安全管理措置のレベルについてAIを活用する事業者側が開示することをもって、「この主体に対してならデータを預けられる」といったことが外形的に判断できるなら、そうした安全管理措置を前提に本人同意の対象からは除外するといった形もあり得る。その部分は事業者自身が挙証するスタイルに変えていくほうがいい。

三点目として、差別的な取扱いにつながる等の問題は確かにある。しかし、統計的なデータ利用の結果が当該個人にフィードバックされるもの、

まさに本人が重大な差別をされ得るような類型は、当然引き続き規律の対象となるべきである。ただ、一点目の論点と被るが、自分の権利利益と直接に影響のないところまで「AIだから」という理由で規律するのはナンセンスなのではないか。

ガバナンスの仕組みについて　「参考資料１－１（参考４）２・３」関連

（事務局）

○　どういったAIモデルの作成であれば、統計化と同じと言えるのかについては、まさに我々も十分な知見があるわけではないが、例えば、基盤モデルの作成等、大丈夫と言える範囲も想定できるのではないかと考えている。一方、ファインチューニング等もいろいろやっている中で、どこまではいいかというところは、もう少し色々と議論していく必要があると思っている。どういうものであれば統計化に近い、統計処理に近い、問題ない処理、モデル作成と言いうるのか、御知見をいただきたい。あと、個人データの場合、例えば第三者提供の同意は不要とするとか、公開情報からだとしても要配慮個人情報の取得・同意を不要とするとしたときに、それが統計処理に近いような、AI モデル作成のためだけに使われるか、それとも実は別の用途にも使われてしまったり、安全管理措置が不十分で漏れてしまったり、問題が起こる可能性があるか、個人の権利利益の侵害につながるか、ということで、つながらないという範囲であれば大丈夫ですと言えるとすると、やはり一定のガバナンスが必要というところもあると思われる。そのガバナンスをどう確保していくのか議論が必要。例えば委託になると、委託はAからBへの委託なので、Aからの業務の範囲内ということとなるが、委託というよりも、第三者提供に近いとしたとき、第三者提供した後に全くガバナンスもないと、別のことに使われてしまう可能性もある。公開の場合、本当に第三者提供した先でAIモデルの作成のためだけに使われるのか、それともよからぬ人が別のことにも使ってしまうのかといった部分等についてもう少し議論が必要ではないか。この人なら信頼が置けるという人と約束を取り交わし、その人だけに提供して、その人との間でこういう目的のためだけに使ってと言った上で、場合によってはプライバシーテックなどを使うのかもしれないが、何らかガバナンスが効いているのであれば、説明しやすいという点をどうお考えか。

（ディープラーニング協会）

○　現状、Common Crawlでデータセットが既に公開をされていて、そこで何か悪用が報告されているかというと、知る限りはない。データの規模やフ

ォーマットからして、これに使うとは考えられないみたいなところはある。例えばそこから個人情報だけ抽出して公開しましたといったことがあった場合、著作権法と同じく、目的規制みたいなところをかけていって、それはもはや統計ないし汎用的なものを出すという目的ではないのではないかという点を規制していくことはあり得るか。

（情報通信研究機構）

○　要するに統計処理をして、そもそもアウトプットでは個人を特定できないから大丈夫という議論だが、よく言うのは、生成AIは勃興期にすぎないので、この技術がどういう方向に進んでいくかよく見えないところがある。そこで統計処理だったらOKとしてしまうと、新しい技術が出てきたときにこれは大丈夫かみたいなことで、またこうしてヒアリングを受ける世界が待っている。実際、著作権法改正で、最初、ウェブデータの収集自体がグレーだった時代に情報解析のためならOKみたいな話になった。最初は情報解析＝統計処理みたいな議論だったが、最新のものではたしかもうすこし間口が広くなっている。技術の進歩に従ってそうなった経緯があり、同じ轍を踏まないように、そういう方向性で検討いただけるとありがたい。

（事務局）

○　おっしゃるとおりで、技術ニュートラルという話もあり、その点は非常に重要。特定のモデルとか、特定の手法だけがOKというやり方よりは、特定の個人に結びつかない、あと、一定のガバナンスが効いているとか、そういうこと要件があったとして、具体的にどういったものが許されるのか、議論した上で、今ならこうだと確認していくやり方もあるかと思っており、そうすると、技術の進展に応じて制度そのものを変えるというより、運用に近いところで調整するというやり方も考えとしてはあり得ると考えられるか。

（AIガバナンス協会）

○　我々としては、技術変化に応じてヒアリングに都度呼ばれるような状況はある種仕方がないことだとも思っている。守らなければならない価値や規範だけが法律等で特定されることが理想であり、それを実現するための方策の部分は技術的に変わっていくものなので、ガイダンスなどに落とした上で、密にマルチステークホルダーで議論して、毎回ある程度の相場観をつくっていくという作業をせざるを得ない。その中で、例えば「こういう事例は許容されない」「このような個人の紐付けは危険である」などといった具体的なケースに即した議論を行っていくべき。要はコアなプリンシプルの部分はしっかりと法で定め、具体的な保護方法のレイ

ヤーについては、都度技術の変化を捉えながら、まさにユーザー側が何を求めているのかという部分も踏まえ、中間規範としてアップデートしていく形が現実的なのではないか。

（プライバシーテック協会）

○　個人の特定などに使わない一つの類型として、統計情報などがある。統計情報はAIモデルに限定しておらず、単に属性と属性の相関関係の情報といったものもある。個人データを突合にして、そういった相関係数だけを出すようなものはいい。AIの学習も高度でないものもある。そういったものは少なくともOKにするということもあるかもしれない。また、システム全体で考えたほうがよく、学習モデルだけでなく、その先の処理で守っている機能があり、トータルとして個人の評価・決定をしているシステムになっていないことをもって、判断しても良いのではないか。

（情報通信研究機構）

○　先ほど一般人の私一人の情報を学習データから抜いても大勢に影響はないと申し上げたが、要するに一般の個人一人の情報を抜いても、出力の大勢にはほぼ影響はない。がらっと変わることはあると思うが、その情報が入っている場合と入っていない場合で、明らかにその情報が入っているがゆえに、例えば、その個人が差別を受けるような出力ばかりが大量に出るということは、恐らくない。ただ、入力で個人名を指定すると、たまたまどこかに書いてあったその個人の情報と差別を紐づけるような話が出てしまう可能性が確率的には高まるだろう。サービス段階の入出力の制御等で、そういったところの規律を入れておけばよいのではないか。ただ、その辺のロジックを一般の皆様が感情的に受け入れられるか、それは恐らく気にされている一番のところかと思うが、そこは声を大にして説明していきたい。一方で、人気商売の方々は、出てくれたほうがいい。そういう方たちは、データ中にそういう記述が多くあるので、そういう出力が出てくる可能性は高まると思うが、それはある意味公人のように受け取られる可能性もあると思うので、あまり問題にはならないかもしれない。一般の人たちの個人情報は、技術的視点で見ると、入出力で氏名等をフィルターする等の措置を講じておけば、リスクは大きくはないのではないかというのが技術者としての感覚。

データについて責任を負う主体について　「参考資料１－１（参考４）３・４」関連

（事務局）

○　三点お伺いしたい。一つ目は、（参考４）１の本人関与について、どう

しても私たちは自分たちのデータを自分たちで管理したい、私たちが納得しない中では使ってほしくないという方々に対して、どういう形であれば理解していただけるか。AIの技術などを分からない方もかなりいるものと思われるが、その場合にはどの程度まで説明ができるか、または、そこまでの説明は必要ないとお考えか、という点。

二つ目は、（参考４）４が近いかと思っているが、個人情報や個人データについて、どの主体がどのように取り扱うのか、また、その前提でどの主体がどのような責任を負うのか、という点。例えば、データを取得した者と、そのデータを処理する者との関係で、どちらがどのような責任を負うことが適切と考えるのか。取得する者は提供した後は、その後の取扱いに責任がないということでよいのか。また、受け取った者はどのような責任を負うべきなのか。データを収集する者と、生成AIのモデルをつくる者と、サービスを提供する者は、それぞれに異なる、という指摘があったが、結局は、現状の個人情報保護法の規定に関しては、特に収集する者に関する要配慮個人情報取得に係る同意、第三者提供に係る同意ばかりが論点となっているように感じる。一方で、皆様の指摘は、むしろ、データを収集する者以外の方も含めた責任を考えるべきということのように感じるが、それぞれの者がどのような責任を負うことが適切だと考えておられるのか、お伺いしたい。

三つ目は、公的な規律ではなく、民間の自主的な取組で対応することを想定しておられるように感じるが、NICTがおっしゃったような、安全保障の部分を本当に民間の自主的な取組に任せて大丈夫なのかという議論もある場合に、公的な規律の議論もあり得る一方で、一般法としての個人情報保護法が対応するのかどうかという議論についてお伺いしたい。法的な規律として、しっかりした公の管理として、例えば次世代医療基盤法みたいなものであれば、政府として事業者を認定する仕組みが必要な場合もある。そこまでやる必要はないということだと思うが、それをどこまで実効的な形で、安心できる形で行わせることを想定されているか。

（情報通信研究機構）

○　三番目の民民というお話について、契約を結んでデータを渡す際には、ここの企業は契約書等に書かれている我々の意図と異なることをするかしないかという観点で判断すべきかと思う。あとは、共同研究の契約書をつくるのに１年かかった。こういう悪用をしてはいかぬとか、こういうリスクはちゃんと対策をしなさいみたいなことがたくさん書いてあって、非常に長い契約書だが、制度がないからつくらざるを得なかった。相当研究時間が削られたような状況。だから、その辺は、国のどこかの部署が引

き取ってやっていただけるなら、我々としてはウェルカム。ただ、それが個人情報保護法の範疇に入るのかどうかは不明。

（AIガバナンス協会）

○　一点目の本人関与について、そもそもAIに学習されること自体の意味も一般的に十分に理解されていない部分もあり、そこは説明していくしかない。もう一つ重要なのは、リスクベースアプローチと言っているが、個人情報保護法の目的にもあるとおり、個人データの活用には便益も必ずあるはずなので、「個人データなどを全て排除したら、どういった便益が損なわれるのか」ということは、NICTのプレゼンにもあったように、リスクとセットで議論していく必要がある。自分に対する直接的なデータに基づく差別などが想定されない場合も含め、あらゆる場面でデータをコントロールすることが本当に求められているのかどうかは、もう少し状況を具体的に特定して、意見を取っていく必要がある。

二点目の責任分担のところは、重要な検討課題。隣接する法領域での責任分担の例を見ると、例えばEUのAIに関わる製造物責任の議論では、従来の製造物責任法と同じように処理している場合もある。一方著作権関係だと、最近大手AIベンダーの中では、自社のAIツールの使用で著作権侵害の訴訟が起きた場合、一定の条件の下で賠償を負う、つまりエンドユーザー側に対しても、上流側が責任を負うという契約にするケースが出てきている。恐らくそうしたAIベンダーの判断の意図としては、学習データは完全にはコントロールしきれないので、そこは賠償の責任を負ってでも、製品を広めるほうが良いという考えがあったのではないか。各主体にどのようなインセンティブと、どの程度のコントロール可能性があるか、という点が考慮されていく必要がある。ただ、ここは実務スタンダードがまだ確立していない領域だと思われ、契約実務の動向も含めて注視していく必要がある。

三点目については、別途お話したこととほぼ同じで、安全保障のような重要な公益に関わる問題については、恐らく他の規律がかかってくるので、民民の契約においても、「法執行のためにせざるを得ない」といった事項に落とし込まれて規律されるのではないか。先ほどのゴールベースの法という観点でいうと、恐らくほかの法領域でカバーされるべき部分だろうと考えている。

（プライバシーテック協会）

○　一点目の本人関与のところ。当協会の提案における突合前の事業者Aと事業者Bは個人データとして持っているので、ここに対しての本人関与はもちろんあるべき。統計目的に関する本人関与は今も必要ない。突合して

統計情報を得るのも同様であるため不要という説明で良いと考える。二点目の提供先、提供元について。法律の専門家ではないため不正確かもしれないが、GDPRには日本のような第三者提供はなく、また、統計目的の本人関与も不要と理解。日本の様に第三者提供したら当初目的が消えることはないので、結論としては、提供先、提供元のどちらかに責任を寄せるのではなく、トータルの処理として見て、両方が責任を負うのが良いのでは。

（ディープラーニング協会）

○　一点目の自分は使ってほしくないという人に対して、どこまでやるかということに関しては、本人関与の在り方を拡大することはありうるとしても、例えば削除請求があったときに、その人の情報を全部削除できるかは別問題である。現状でも、例えばAIで顔画像を解析しますといったときに、自分の顔が映っているから削除してくれと言われたとき、結局、事業者は顔を持っておらず、顔を数値化した特徴量データを持っているに過ぎないので、削除してほしいのだったら顔写真を何枚かください、それで特徴量データと突合して、特徴量を削除しますみたいなことをやっている。ただ、そうであっても本当に削除できるか分からないところはどうしても残り、データセットも同じ。どこまでできるかという点は結構難しいので、本人関与の在り方を拡大するとしても、結果責任を課すような形になるのは望ましくない。あとは、統計化のところでいうと、現状は統計化なら、本人は何と言おうがやっていいということになっているので、本当にAIと統計処理が同じならば、理解を促進していくほかない。一点目の誰がどうやって処理をして責任を負うのかという点については、莫大なデータベースの中に散在している話であれば、現時点でもそもそも提供に当たっての規律はなくて、取得した各事業者がどう使うかに規律が及んでいるところだと思い、そういう形でいいのではないか。民民でよいのかという点については、安全保障などの論点については民民ではできないところが出てくると思うが、こういった論点は個人情報保護法の範囲は超えているように思われる。個人情報の関連でいうと、確かに匿名加工などだと、おっしゃったとおり、次世代医療基盤法で認定をするところもあるかと思うが、AIの具体的な技術等について官が認定するのは難しいので、そこはある程度民民でやらないと、硬直的になり過ぎてしまうのではないか。

（情報通信研究機構）

○　一番目と二番目について回答する。一番目について、要するにAIの学習でなくても、例えば大量のウェブページから高所得者のリストをつくる

ことはできてしまう。それはある意味AIの学習の問題とは別で違法で規律をかけるほうが望ましい。当然AI学習目的で大量のデータを渡されたときに、そこから高所得者リストで強盗に入るターゲットリストをつくったとなったら、それはアウトとする規制はあっていいのではないか。

学習したAIが結果として個人に悪影響を与えるようなケースはありえるとは思うが、例えば入力と出力にフィルターをかける、場合によっては本人から申請があれば、本人の関連情報は一切出さないようにするとか、そこもベストエフォート的な話があるが、そういったところで努力目標を課す的な、そういう対応策もあるのかもしれない。一方で、学習自体で私の微々たるデータがあり、モデル全体に影響が出ることは極めて少ないのではないかという点は声を大にして宣伝していかないといけない話。

あとは、二番目の責任分担について、データ提供を受けた側が何をするか、もうちょっと真剣に規律を考える必要があるのではないか。データ提供を受けなくても、自分でウェブページを収集してきて、生成AIの学習とは全く別の方法で悪いことをしようと思ったらいくらでもできるので、AI学習とか、データ提供と一旦切り離し、これはウェブデータを使った悪いことだと書き下していくことはできそうに思うが、いかがか。当然その過程で、生成AIに個人の情報を出させて、それで名簿を作ると言った話も含めて、生成AIを使った悪用の仕方もそこに書かれるべきかもしれない。最も今の生成AIでは正しい個人の情報は特に一般人の場合、滅多に出てこないと思うが。生成AIの学習データの提供、取得と、本来悪いとみなされる行為は、独立なのではないか。

以上