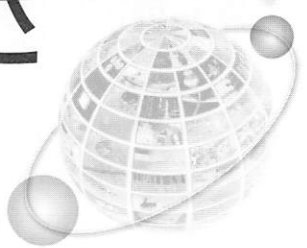
連載262 市民と行政を結ぶ情報公開・プライバシー保護

対話型AIの危うさ

情報公開クリアリングハウス理事 奥津 茂樹



対話型AIのChatGPTが「すごい！」らしい。22年11月に公開されてから数か月だが、社会的な注目が一気に高まっている。最先端のAI技術を活用して、利用者の問いかけに対する回答を瞬時に繰り出すという。新しもの好きの私は早速アカウントを取得して、遊び半分で利用を始めた。回答の文章は読みやすく、あらゆる質問に回答する能力の高さに驚かされる。いずれは本連載も取って代わられるのだろうかなどと妄想しつつも、開発途上ゆえの課題が気になり始めた。

エゴサーチからの学び

すでに課題に対する指摘は各分野から出始めている。日本の大学の中には、学生による使用に注意喚起をした例もある。また、23年3月、イタリアは個人情報保護を理由に一時的な使用禁止を打ち出した。

新しい技術に対するクリティカルな反応は、社会の健全性の証明である。しかし、日本社会は時代の変化・風潮への同調傾向が強く、「すごい！」という好意的な反応が多くなりやすい。そこで、今回は個人情報保護面での課題を指摘し、自治体現場での慎重な対応を求めることとした。

その起点となるのが、私自身のエゴサーチである。エゴサーチとは自分の名前をキーワードにして、ネットやSNSでの検索をすることをいう。初めてのChatGPTの利用は、こんな質問を書き込んでみた。

「奥津茂樹について教えてください」…さすが最先端のAIである、質問に対する記述の生成（generating）が直ちに始まった。

「奥津茂樹（おくつしげき）は、日本のプロ野球選手であり、現役時代のポジションは投手です。1971年に広島東洋カープに入り…」

以下、「プロ野球選手」としての記述・説明が続く。念のため同姓同名の選手がいたのかを調べてみたが、そうした事実はない。なぜ、こんな内容になったのかは不明だが、記述全てが不正確な個人情報であった。

ただ、対話型ゆえに漠然としたオープンクエスチョンだと、こうした不正確な情報を引き出しやすい。そこで、私の名前の前に「情報公開クリアリングハウス理事の」という肩書きをつけると、さすがに「プロ野球選手」という記述はなくなった。しかし、後述するように、より問題が大きい記述が新たに加わった。

エゴサーチだから笑い話で済ませることができる。しかし、これが他者による検索だったらと思うと笑ってはいられない。閲覧だけならともかく、不正確な記述がSNSに貼り付けられ、データ・文書に転記される可能性は誰についてもある。

不正確な個人情報が生成され、その利用・拡大が広がると権利侵害が深刻になる。エゴサーチから学んだのは、ChatGPTによるプライバシー侵害の危うさである。

正確性の確保

個人情報保護法22条は「利用目的の達成に必要な範囲内において、個人データを正確かつ最新の内容に保つ」と規定し、個人情報取扱事業者にデータ内容の正確性の確保を求めている。ちなみに21年改正により、個人情報取扱事業者には国の機関や自治体が含まれることになった。

ただし、正確性が求められるのは

＊事実証明のための
画像を公開しています

「個人情報」ではなく、「個人データ」であることに留意したい。これは「個人情報データベース等を構成する個人情報」と定義されている（同法16条3項）。そのため、対話型AIの回答を取得しデータベース化しなければ、個人情報の正確性の欠如は問題にならないと思われる。

なお、国の「個人情報の保護に関する法律についてのガイドライン（通則編）」（個人情報保護委員会16年11月策定）は、同法20条1項の適正取得の説明の中で、「取得」について以下のように説明している。

「個人情報を含む情報がインターネット等により公にされている場合であって、単にこれを閲覧するにすぎず、転記等を行わない場合は、個人情報を取得しているとは解されない」

以上のことから、私が試したエゴサーチのように対話型AIによって特定個人の情報を生成させることは、たとえ内容に誤りがあっても、直ちに個人情報保護法に違反するとは言えない。ただし、「転記等」によりデータベース化した場合は、個人情報に誤りがあることが問題になる。

適正な取扱い

対話型AIによって生成される個人情報が不正確であることは、本人にとって不快で不安である。ただ、上記のように、一見すると個人情報保護法上の問題がないように思われる。本当にそうなのだろうか？

同法3条は「個人情報は、個人の人格尊重の理念の下に慎重に取り扱われるべきものであることに鑑み、その適正な取扱いが図られなければならない。」と定めている。

私の経験を素材として示したが、事実に反する個人情報の生成は、「個人の人格尊重」の対極にある事態である。そして、これを放任しておくことは、決して「適正な取扱い」とは言えない。

確かに対話型AIは開発・発展途上であり、自己学習能力が高いため、いずれは正確性の高い個人情報を生成し得る。しかし、だからと言って、個人の人格を侵害する事態を放置し、適正な取扱いを怠るべきではない。

これを痛感したのは、私自身のエゴサーチの中に含まれていた。以下の記述である。

「ただし、奥津氏は元電通の役員として過労自殺問題に関与したことがあり、一部の人からはその過去を問題視する声もあることは事実です。しかし、ICHは奥津氏を理事として採用する際に、彼の過去を踏まえたうえで採用を決めたとされています」

念のために申し添えておくが、私は元電通の役員ではない。また、同社で過労自殺問題があったことは事実だが、私がそれに関与した事実はない。なお、ICHとは情報公開クリアリングハウスの英語表記（ICJ）の誤りである。

ChatGPTにはこうした不正確な記述の再考を求めるRegenerate responseというボタンがある。これをタップすると以下の記述が生成された。

「奥津氏が電通の幹部として関わった問題については、団体の活動に影響を与えないようにするための議論が行われている」

責任主体不在の制御不能

もはや笑うしかない。事実無根の個人情報が生成され続けることは、私の人格を侵害している。さらに、記述に取り上げられた企業やNPOの信用や名誉を毀損している。

わからなければわからないと言えばいい。答えなくても構わないこと、答えてはならないことなどの分別がつかない。それは、対話型AIが制御不能に陥っていることを象徴している。

不正確であるだけでなく、正確性を高めていくことにも、対話型AIの危うさがある。たとえば前科のような社会的差別の原因となり得る情報が、簡単に生成されてしまう。私のように不正確ならば笑えるが、正確だからこそ笑えないことがある。

個人情報保護法が「適正な取扱い」の一つとして重きを置くのが、「要配慮個人情報」である。これについて同法2条は、以下のように定める。

「本人の人種、信条、社会的身分、病歴、犯罪の経歴、犯罪により害を被った事実その他本人に対する不当な差別、偏見その他の不利益が生じないようにその取扱いに特に配慮を要する」

私について対話型AIが生成した記述には上記が含まれている。これについて、利用目的をはじめとする何の説明もなかった。そして、本人同意の手続きもない。しかも、こうした法令違反に対して誰が責任を負うのか不明である。制御不能な無法状態をどうしたら良いのだろうか？