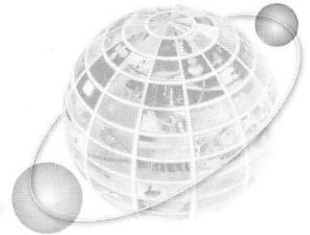


AI のジレンマ

情報公開クリアリングハウス理事 奥津 茂樹



本 連載262回で「対話型AIの危うさ」を取り上げた。その進化は急速で、23年5月にはGoogleがGeminiをリリースした。今後も新規参入が相次ぐと思われる。また、自治体現場における関心も高まり、現場での利用を始めたところもある。また、ChatGPTを含む総称も、あつという間に変更されつつある。現段階におけるネット検索では、対話型AIではなく生成AIと表現する例が圧倒的に多い。これもまた急速な変化を象徴している。課題も次々に現れ、急速に改善されていくと思われる。

ChatGPTの変化

急速な改善として、追記しておきたいのが262回で取り上げたエゴサーチである。

そのとき、ChatGPTが私について「プロ野球選手」と記述・説明したことを紹介した。同姓同名の選手は実在せず、まったくのデタラメである。同様のフェイクニュースの生成は、著名な方々によるエゴサーチでも起きているという。

東大の憲法学者は「フィギュアスケートの選手」とされた(朝日新聞23年5月25日)。ある県知事は「出身大学や経歴」に大きな誤りがあった(産経新聞23年5月23日)。私のような無名人物ならともかく、著名な人物までもデタラメな記述・説

明だったことに改めて驚く。

このように正確性への疑念が高まったからかもしれないが、最近、変化が見られる。先日、改めてエゴサーチをしたところ、以下のような回答だった。

「残念ながら、私のデータベースには奥津茂樹という具体的な人物に関する情報は見つかりませんでした。奥津茂樹について具体的な情報や背景があれば、それに基づいて詳細な情報を提供することができますかもしれません。その他の情報があれば教えてください。」

わからないのに無理に答えようとすると、フェイクという傷口を広げていく。そんな前非をAIは悔いたのだろうか。「わからない」と回答する知恵を身につけたと思われる。

さらに、前科という要配慮個人情報

報のデタラメさも以下のように修正された。

「私は個別の個人の前科情報を提供することはできません。個人の前科情報は個人のプライバシーに関わる機密情報であり、法的な制約があるため、一般の人々がアクセスできる情報ではありません。」

「過労死自殺問題への関与」や「覚醒剤取締法違反」など、ありもしない私の前科を次々に生成した数か月前と異なり、大きく改善された点の一つである。

なお、前科とともに典型的な要配慮個人情報である病歴のエゴサーチをすると、現在のChatGPTは以下のように回答する。

「私は奥津茂樹さんの個別の病歴について情報を持っていません。また、個人の病歴は医療情報の一部であり、プライバシーの保護が重要な要素となります。」

「不知」という回答が正確性を向上させたわけではない。しかし、「個人の人格尊重」という個人情報保護法の基本理念に反する事態は回避された。

正確性と汎用性



なお、個人情報以外の面でも正確性が疑われる事例がある。たとえば、本誌についてたずねると、ChatGPTから以下のような回答があった。

「月刊ガバナンスは、企業のガバナンス（統治）に関する情報を提供する日本の雑誌です。経営者や企業経営に関わる方々を対象に、組織の統治やリスク管理、法的規制、CSR（企業の社会的責任）など、経営における重要なトピックやテーマを取り上げています。」

特定のキーワード（ガバナンス）に引きずられて、誤った認識、理解を次々に繰り出す本質は、何も変わっていないように思える。他の情報との比較をしたり、全体の文脈を考慮した総合的な認識、理解はまだまだ開発途上のように思える。

なお、Bardは当初からエゴサーチについては「不知」の回答をしていた。また、Googleの情報検索を活用・反映しているのか、ChatGPTに比べると正確性が高い印象を受ける。実際に、本誌に関わる回答は以下の通りだった。

「月刊ガバナンスは、ぎょうせい1973年から発行している地方自治に関する専門誌です。…自治体経

営に関する情報の宝庫であり、自治体関係者に欠かせない情報源となっています。」

発行年が残念だが、最後の記述のように正しく記述されている。

以上のことから、生成AIの正確性はまだまだ開発途上である。それにもかかわらず、一部の自治体が活用に前のめりなのが気になる。

私のスタンスは活用に「反対」ではなく「慎重」である。そして、「できる／できない」や「望ましい／望ましくない」など、一方の見方に偏らず、常に批判的考察が欠かせないと考えている。

自治体現場で生成AIの汎用性を高めていくには、何よりも正確性の確保が大切だ。連日のように報道されてきたマイナンバーカードの混乱は、性急で前のめりすぎる国の姿勢がもたらした。

正確性を欠くと社会に不信感が広がり、技術の活用を困難にして、信頼回復のためのコストを高めていく。生成AIについても、その悪循環を断ち切らなければならない。

ビッグデータ

AIの正確性を担保するのがビッ

グデータだと説明される。このビッグデータとは具体的にどのような情報を指すのだろうか。そして、本当にビッグデータなのだろうか。そのような疑問を持ち続けてきた。

そして、生成AIを試用しているところ、ビッグデータには限界があることがわかる。その根拠となる記述が、ChatGPTやBardには散見される。

前述した「不知」という回答は、情報が「不存在」であることを意味する。ビッグデータと言われるが、ありとあらゆる情報を取得・保有しているわけではない。

たとえば、前科は、自治体を含めて関係する行政機関が保有している。しかし、それらは厳正に管理され、生成AIが容易にアクセスし、引き出せるような情報ではない。

もちろん例外的な事例はある。たとえば、ネットニュースやこれを引用したSNSの書き込みを、生成AIが拾い上げ提供する可能性がある。ただ、生成AI提供者が前科情報の生成・提供禁止をルール化すれば、前科情報の流出リスクは低減する。

また、個人の病歴も本来であればビッグデータには存在しない。たとえばレセプト（診療報酬明細書）は

保険者が、また、カルテ等の医療個人情報情報は病院等の医療機関が、厳正に管理しているからだ。ここでも、病歴情報の生成・提供禁止というルール化が重要になる。

深く考察していない直感的な仮説だが、前科や病歴等の要配慮個人情報の取得は禁止しなければならぬ。主体がないAIが同意を得ることは困難で、本人の関与による制御は望めないからだ。なお、取得禁止が容易でないときは、前述の改善のように「不知」という形で、生成・提供を禁止しても良い。

もう一つ重要なのは、さまざまな情報をビッグデータ化できない現状を維持・強化することである。セグメント（分割）ごとの厳正な管理を徹底し、外部への流出を少なくすることで、生成AIによる要配慮個人情報の提供リスクは低減する。

言い換えるならば、生成AIは共有可能な限られた意味での「ビッグデータ」を活用する。そうした制約を課すことで個人情報保護が図られ、他の情報についても過度な依存が抑制できるのではないか。

生成AIにはそれなりのデータと正確性しかない。という醒めた見方が、保護と利用の適正化を担保する。