



第3回：マスキングの先にあったもの

本音を言えば、生成 AI はもっと気軽に、自分のために楽しく使うものだとは私思っている。

しかし、あまりにも文章が上手なので、福祉の現場でも使いたいと思う人が多いのは当然だと思う。

前回、福祉関係者が生成 AI を使おうとしたとき、必ずと言ってよいほど個人情報の問題にぶつかることを書いた。そして、その対策の一つとして、私自身がマスキングアプリを開発していることにも触れた。

私が作っていたのは、生成 AI に文章を渡す前に、個人を特定できそうな情報を自動的に置き換えるアプリである。

氏名なら「氏名 1」、住所なら「住所 1」、事業所名なら「事業所 1」。医師の名前や担当者名、親族の名前なども、できるだけ固有名詞が残らないように置き換える。

その状態で生成 AI に文章を読ませ、記録を整理したり、文章を整えたりする。

少し手間はかかるが、個人情報をそのまま生成 AI に渡すよりは安全だろう。

私はそう考えていた。

開発は比較的順調に進んでいた。

そこであるとき、このマスキングがどの程度有効なのか、試してみることにした。

架空の高齢者の支援記録を作った。もちろん実在する人の記録ではない。氏名や住所、機関名などはマスキングした。一方で、その人の発言や行動、家族との関係、生活の変化などは残した。

それを生成 AI に読ませてみた。

いつ頃から認知機能の低下が見え始めているのか。

家族との関係にどのような変化があるのか。親族によるネグレクトを疑うとすれば、どのあたりから兆候が見えるのか。

複数の記録を時系列に並べて、いわば定点観測のように分析させてみたのである。

結果は、想像以上に良好だった。

氏名も住所も伏せているのに、生成 AI は記録の変化を追い、かなり自然に特徴を整理した。

私は感心しつつも、その画面を見ながら、途中でちょっと違うことを考えた。

固有名詞をこれだけ消しているのに、なぜ AI はここまでその人について分析できるのだろう？

裏を返せば、氏名や住所を消したところで、その人についての情報は文章の中にかなり残っているということではないか。

発言の内容、家族構成、これまでの経過、疾病、生活状況、支援者とのやり取り。

そうしたものが組み合わさって、その人らしさを形づくっている。

私はそこで初めて、マスキングというものを少し違う角度から見ようになった。

なぜ、アプリを作る前にこのことに気づかなかったのだろう。

そこで、個人情報保護法の個人情報の定義を、あらためて読み直した。

そこには、氏名や生年月日だけではなく、「その他の記述等」によって特定の個人を識別できるもの、さらに、他の情報と容易に照合することによって特定の個人を識別できるものも含まれると書かれている。その文章を読みながら、私はだんだん困ってきた。

生成 AI の研修などでは、ときどき、

「個人情報を使う場合は、本人が特定されないようにマスキングしてください」

という説明を聞く。

私自身も、その考えをかなり素直に受け取っていた。

だからマスキングアプリを作った。

けれども、では、どこまでマスキングすれば「本人が特定されない」と言えるのだろう。

氏名を消すだけでは足りない。

住所、勤務先、病院名、家族の名前も消す。

それでも、地域や年齢、家族構成、生活歴、独特のエピソードが残っていれば、身近な人が読めば「あの人のことではないか」と分かることがあるかもしれない。

では、それも消せばよいのだろうか？

けれども、情報を消せば消すほど、生成 AI に読ませる意味も薄くなっていく。

私が試したように、認知機能の変化を見つけないのであれば、本人の発言や行動の変化は必要である。

家族によるネグレクトの可能性を考えたいのであれば、家族との関係や、それまでの経過も必要になる。

具体的な背景があるからこそ、生成 AI はそれなりの分析ができる。

しかし、その具体性こそが、その人をその人たらしめる情報でもある。

ここまで考えたところで、私は完全に立ち止まってしまった。

どこまでマスキングすればよいのか。

その境界線を、いったい誰が判断するのだろう。

ある人には分からなくても、本人を知る人なら「あの人のことではないか」と気づくかもしれない。

しかも福祉の現場で、記録を AI に入れるたびに、

「ここは残してよいだろうか」

「これは消した方がよいだろうか」

「このくらい抽象化すれば大丈夫だろうか」

と考える運用が、本当に続くだろうか。

そんなことを考えているほど、現場は暇ではない。

私はだんだん、これはマスキングアプリの精度を上げれば解決する問題ではないのではないかと思うようになった。

同じころ、何度も読み返していた文章がある。

2023年6月に個人情報保護委員会が出した「生成 AI サービスの利用に関する注意喚起等について」である。

そこには、本人の同意を得ずに個人データを含むプロンプトを入力し、その情報が「当該プロンプトに対する応答結果の出力以外の目的」で取り扱われる場合には、個人情報保護法に違反する可能性があること、そのため生成 AI サービスを提供する事業者が、個人データを機械学習に利用しないことなどを十分確認するように、という趣旨が書かれている。

この文章に、私はかなり長い間悩まされた。「機械学習に利用しないことを確認する」とは、学習に使われない生成 AI なら個人データを入力してよいという意味なのだろうか。



私は長くそう考えていたが、やがて一つの疑問に行き着いた。

AI が情報を学習に使わないことと、その個人情報を AI に入力してよいことは、同じな

のだろうか。

どうも違う。

そこにたどり着くまで、私自身、数か月かかった。

おそらく私は、「仕事で個人情報を生成 AI に使えるようにしたい」という気持ちをどこかに持ちながら、法律を読んでいたのだと思う。

人は、自分がそうであってほしいと思う方向に文章を解釈したくなる。

私も例外ではない。

けれども、

「学習には使われません」

という説明だけで、本人の許可なく、その人の病気や家族関係や生活上の困りごとを生成 AI に渡してよいと考えるのは、私には難しかった。

もし本人から、

「私があなたに話したことを AI に読ませたのですか」

と尋ねられたら、どう答えるだろう。

「大丈夫です。AI の学習には使われませんから」

という説明で、本人は納得するのだろうか。

さらに、学習利用だけが問題ではない。

どれほど安全性の高いサービスでも、漏洩の可能性を完全にゼロにすることはできない。

サイバー攻撃もあれば、アカウントが乗っ取られることもある。人間側の操作ミスもある。

もし何か事故が起きたとき、

「そもそも、その人の情報を AI に入れないで起きなかったのではないかな」

と言われたらどうするのだろうか。

その問いは、最終的には入力した側に返っ

てくる。

そんなことを考えながら生成 AI を使っていると、もう一つ気になることが出てきた。生成 AI は、とてもそれらしい文章を書く。私は日常的に生成 AI を使っているが、元の情報にはなかったことが、もっともらしい文章として紛れ込む場面を何度も見てきた。背景情報が足りないとき、生成 AI は一般論や推測で文章を補うことがある。

私はこの現象を見るたびに、生身の人の処遇や悩みを扱うことに、生成 AI はあまり向いていないのではないかと思うに至った。そもそも支援記録に書かれているのは、その人の生活のほんの一部である。支援者が見聞きした断片を AI に渡しているにすぎない。

AI の答えが驚くほど自分の感覚に合うこともあるが、それを「AI がその人を理解した」と考えない方がよい。限られた情報から返された一つの可能性として受け取るくらいの距離が必要なのだと思う。

だから私は今回、一つの補助線を引き直すことにした。

個人情報をどうにかマスキングして生成 AI に入れようとするのを、一度やめる。そもそも個人情報を生成 AI に入れるという発想を持たないところから考え直すことにした。

かなり悩ましい補助線ではある。

実際、現場には AI にケースのことを相談したいというニーズがある。

自分の支援が本当にこれでよいのか、別の視点から意見がほしい。

ケアマネジャーや相談援助職の方から、

「個人情報を入れたらだめなんですか」

と、がっかりしたように尋ねられることもある。

人の生活について考え、支援の方向を決める仕事は重い。誰かに一緒に考えてほしいと思うことは当然である。

だから、そのニーズ自体を否定したいわけではない。

ただ少なくとも、「私たちの組織では、この条件で個人情報を生成 AI に入力してよいことにしています」「その根拠はこれです」「本人にはこのように説明します」

と明確に言えないのであれば、私は入力しない方がよいと思うようになった。

マスキングアプリの開発に取り掛かり、そして立ち止まったことで、私は一つ手前の問いに戻るようになった。

補助線は、正解ではない。

考えるために、一度引いてみる線である。だから、考えている途中で引き直すことがあってもよいのだと思う。

今回はずいぶん遠回りをした。

マスキングをどうするかという問いから始まって、どこまで消せば個人情報ではなくなるのかと考え、その境界線が簡単には引けないことに気づき、最後には、「どうやって AI に入れるか」ではなく、「そもそも、AI に入れる必要があるのか」というところまで戻ってきた。

これが、マスキングアプリを作りながら私がたどり着いた、今回の伏線回収である。