AI法律相談室

第2回 油断禁物…機械学習用データの「個人情報」

大本総合法律事務所 弁護士：関原（せきはら） 秀行（ひでゆき）

機械学習には，大量の学習用データが必要な場面が少なくありません．機械学習に利用されるデータを法的側面から分類すると，「個人情報」を含むデータと「個人情報」を含まないデータの2つに大別できます．

日本では，個人情報保護法（個人情報の保護に関する法律）が存在し，「個人情報」を含むデータについては，一定のルールに従って取り扱うことが求められています．

学習用データの収集前にルール違反の恐れがあることに気づけば問題ありません．学習用データを収集していざ機械学習を行う段階でルール違反に気づいた場合には，手間をかけて収集したデータ自体が利用できないことにもなりかねません．そのような事態を避けるために，「個人情報」を含む学習用データ利用時の主なチェック・ポイントを一緒に見ていきましょう．

● 1…顔や氏名だけでなくIDも規制の対象に

個人情報保護法の規制対象は「個人情報」です．そこで第1に，利用予定の学習用データに「個人情報」が含まれているかどうかを確認します．

個人情報保護法は「個人情報」につき，生存する個人に関する情報であって，次の①または②のいずれかに該当するものと定義しています．

①…顔認識データ，指紋データ，マイナンバーなどの個人識別符号が含まれるもの
②…氏名などのように特定の個人が識別可能なもの

確認の手順としては，①が学習用データに含まれているかどうかを確認した上で，含まれていない場合には，②が含まれていないかを確認することになります．

②の特定の個人が識別可能な情報には，その情報単体では個人を識別できない情報であっても，他の情報と容易に照合でき，それにより誰の情報かが分かるものは「個人情報」に含まれるとされています．そのため「購買履歴」を学習用データとして利用するために「氏名」を削除したとしても，「購買履歴」が「ID」によって「氏名」とひもづいていれば，IDを介して購買履歴と氏名を照合することによって誰の購買履歴かが容易に分かるため，これらの情報は全て「個人情報」に該当し，個人情報保護法のルールに従い取り扱う必要があります．

● 2…データの取得時には所定の手続きを踏む

犯罪行為によって学習用データを取得することは当然許されません．それに加えて氏名などの「個人情報」を含む学習用データを取得する際には，原則として事前に「個人情報」の利用目的を特定した上で，その利用目的を個人情報の対象となる本人に通知・公表する必要があります．

具体的には，プライバシ・ポリシに個人情報の利用目的を記載することによって利用目的を特定した上，そのプライバシ・ポリシを自社のウェブ・サイトに掲載することにより公表する例がよく見受けられます．

また，人種や健康診断結果などのセンシティブ・データは，原則として，本人の同意を得なければ取得できません．

● 3…目的の範囲内で利用する

個人情報保護法は，利用目的の達成に必要な範囲を超えて「個人情報」を取り扱ってはならないと定めています．そのため，利用予定の学習用データを取得する際に定めた利用目的の範囲を超えて，学習用データを機械学習に利用することはできません．

利用目的の範囲を超えて「個人情報」を含む学習用データを利用する場合には，本人の同意を得るか，個人情報保護法が定める要件に従い，利用目的の変更や匿名加工を行う必要があります．

利用価値があるデータを大量に集めたとしても，利用できなければ意味がありませんので，機械学習に利用予定のデータに「個人情報」が含まれる可能性がある場合には，データの取得時から最終的な利用方法をイメージし，その利用目的を取得時から特定・公表することが重要です．

● 4…データは安全に管理する

「個人情報」は外部に漏（ろう）えいした場合，情報を漏えいされた本人から損害賠償請求を受けるリスクがあるとともに，自社のレピュテーションが低下するリスクもあります．そのため「個人情報」を含む学習用データは外部に漏えいしないよう安全に管理する必要があります．個人情報保護法においても，顧客データベースを構成する顧客氏名のような「個人データ」につき，情報が漏えいしないよう安全に管理する措置を講じる義務を課しています．

● 実際に利用するときは専門家の知恵がほしい

実際の利活用に当たっては専門家に相談するか，個人情報保護法を所管する個人情報保護委員会のウェブ・サイト（https://www.ppc.go.jp/）に掲載されている資料を参照することをお勧めします．

第1回 完全自動運転車が起こす事故はだれの責任か（2018年2月号）