Privacy-Forscher im Interview

„Deanonymisierung geht enorm schnell“

Deutschland soll wieder Spitze in Medizin und Forschung werden, erklärt Bundesgesundheitsminister Karl Lauterbach bei vielen seiner Auftritte. Dafür sollen viele Daten fließen. Doch für den Umgang mit besonders schützenswerten Patientendaten bedarf es auch besonderer Schutzmethoden. Daran arbeitet Prof. Esfandiar Mohammadi mit seinem Forschungsprojekt AnoMed. Die Sekundärnutzung der Daten für die medizinische Forschung stelle uns vor komplett neue Anonymisierungsherausforderungen, sagt er im Gespräch mit dem änd. Denn perfekte Anonymisierung sei unmöglich, ohne alle Information zu zerstören.

©Privat

„Ich bin mir nicht sicher, ob wir uns als Gesellschaft bewusst sind, welchem Risiko wir da gerade entgegengehen", sagt Prof. Esfandiar Mohammadi.

Haben Sie als jemand, der sich wirklich jeden Tag damit beschäftigt, den Eindruck, dass der Bevölkerung die Risiken der digitalen Welt wirklich bewusst sind?

Für eine qualifizierte Aussage wären natürlich repräsentative Studien notwendig. In meinem näheren Umfeld scheinen die Risiken der digitalen Welt eine untergeordnete Rolle zu spielen, falls man nicht gerade Opfer eines Cybervorfalls ist. In der öffentlichen Diskussion beobachte ich, dass IT-Sicherheit zwar ein wichtiges Thema ist, aber der Wille dafür etwas zu tun, ist begrenzt. Ein Grund dafür könnte die mangelnde Berichterstattung sein. Außer in Fachzeitschriften sehe ich nur Berichterstattung über die schwerwiegendsten Cybervorfälle, obwohl in großer Regelmäßigkeit signifikante Cybervorfälle bekannt werden.

Was sollte aus Ihrer Sicht stärker in der Öffentlichkeit diskutiert werden?

Das Teilen von personenbezogenen Daten eröffnet viele Chancen, die wir nutzen sollten. Mir fehlt in der öffentlichen Debatte sowohl eine Diskussion der Chancen, die eine Nutzung von personenbezogenen Daten mit sich bringen kann, als auch eine tiefergehende Diskussion über Deanonymisierungsgefahren.

Konkret würde ich mir wünschen, dass mehr über das Verknüpfungspotenzial verschiedener Statistiken über die gleiche Personengruppe diskutiert werden würde. Ich gebe dazu ein Beispiel: Jüngste Angriffe auf den US-Zensus von 2010 haben gezeigt, dass die Veröffentlichung vieler Statistiken über die gleiche Personengruppe unter Zuhilfenahme öffentlich verfügbarer Informationen genutzt werden kann, um Individuen aus dieser Personengruppe zu deanonymisieren.

Bedarf es denn im Umgang mit den als besonders schützenswert geltenden Patientendaten aus Ihrer Sicht besonderer Schutzmethoden?

Ja, absolut. Es gibt inzwischen auch viele Methoden aus der IT-Sicherheit, die man nutzen kann. Außerdem ist eine aktive Begleitforschung gefragt, die sich den Fragen nach neuen Schwachstellen widmet und untersucht, wie man die existierenden Systeme angreifen könnte, damit sie anschließend nachgebessert werden können.

Aus IT-Sicherheitsperspektive gehen wir ein großes Risiko ein, wenn wir die Daten zentral speichern. Es sollen außerdem europäische Datenräume entstehen, die fordern, dass auf ihnen durchgeführte Studien veröffentlicht werden. Das wird das Gefahrenpotenzial durch die Verknüpfung von Statistiken verstärken.

Ein prominentes Beispiel hierfür ist der europäische Gesundheitsdatenraum EHDS, für den gesetzlich geregelt ist, dass medizinische Studien und Untersuchungen, die im Interesse des Gemeinwohls sind, auf allen Patientendaten europäischer Bürger:innen durchgeführt werden können sollen. Solche Studien sollen laut Gesetzesentwurf veröffentlicht werden. Damit würden ähnlich wie bei einem Zensus über die Zeit etwas ähnliches wie Statistiken über die gleiche Personengruppe veröffentlicht werden. Bei einer klassischen Erstellung von Studien könnten also mit der Zeit ähnliche Angriffe wie auf den 2010er US-Zensus möglich werden.

Gibt es konkrete technische und organisatorische Maßnahmen aus der IT-Sicherheit, die notwendig sind, um die Digitalisierung des Gesundheitswesens sicher und verantwortungsvoll zu gestalten?

Es gibt davon eine Reihe. Die Herausforderungen, die sich aus der Perspektive der IT-Sicherheit stellen, fächern sich mindestens in zwei Teile auf: Erstens sollten wir selbst für die Digitalisierung der Versorgung – der sogenannten Primärnutzung der Patientendaten—sicherstellen, dass die zentralen Systeme und die Organisationen, die alle Patientendaten speichern, im klassischen Sinne der IT-Sicherheit Angriffe verhindern.

Die Sekundärnutzung der Daten für die medizinische Forschung stellt uns komplett neue Anonymisierungsherausforderungen. Dieses Thema liegt zentral in meiner Forschung. Um sicherzustellen, dass in der Zukunft keine der vorher genannten Verknüpfungsangriffe, wie auf den 2010er US-Zensus, durch veröffentlichte Studienergebnisse möglich sind, muss sichergestellt werden, dass der Einfluss jedes einzelnen Patienten auf die veröffentlichte Studie beschränkt wird.

Das ist natürlich nur in den Fällen möglich, in denen genügend Daten vorliegen. Die Forschung der letzten 20 Jahre hat gezeigt, dass die Nutzung von aggregierter Information, z.B. von Statistiken, in Kombination mit randomisierten Methoden, z.B. das zufällige Verrauschen, hilfreich sein können, um Verknüpfungsangriffe und andere Arten von Deanonymisierungstechniken zu verhindern. Ein Lichtblick ist, dass diese Art von Methoden auch von maschinellen Lernverfahren (sogenannten KI-Techniken) verwendet werden und somit eine starke Synergie dieser Forschungsbereiche besteht. Diese Forschungsrichtung ist noch sehr jung, macht aber große Fortschritte. Ich erwarte, dass wir in den nächsten Jahren besser verstehen, für welche Anwendungen wir personenbezogene Daten gut nutzen und schützen können.

Sie sprachen Ihre Forschung an. Das Projekts Anomed, das Sie leiten, ist eines von fünf Kompetenzclustern. Es soll als Katalysator für die Anonymisierungsforschung in medizinischen Anwendungen dienen und die medizinischen Anwender über die Gefahren der Deanonymisierung aufklären. Mit anderen Worten, Sie beschäftigen sich mit den Gefahren, die es gibt, und wie wir uns davor schützen können. Wie kann eine Deanonymisierung passieren und wie real ist diese Gefahr?

Bei der Deanonymisierung gibt es eine noch größere Diskrepanz zwischen dem, was die Menschen erwarten, und dem, was tatsächlich möglich ist. Vielen ist es nicht bekannt, dass Deanonymisierung enorm schnell geht. Um das zu veranschaulichen, nenne ich gerne die App Akinator, bei der es im Prinzip um Prominentenraten ging. Die App hat Fragen gestellt, die man beantworten musste. Nach ungefähr 20 Fragen wusste die App, welche Person gesucht war. Das wirkt dann immer wie Zauberei.

Aber die Mathematik, die dahintersteckt, ist die gleiche, die bei der Deanonymisierung zu Problemen führt. Mit jeder Frage, die die App stellt, kann sie nämlich die Anzahl der möglichen Kandidaten ungefähr halbieren. Und wenn man 20 Mal halbiert, dann hat man ein Millionstel. Nach 33 Fragen ist es ein Achtmilliardenstel und das sind dann alle Menschen auf der Welt.

Genauso ist es, wenn wir Attribute von Menschen veröffentlichen. Wenn wir 33 Ja- und Nein-Fragen über Personen stellen, ist eine Person mit hoher Wahrscheinlichkeit eindeutig identifizierbar. Vielleicht kennen wir den Namen nicht, aber wir haben die Person mit ungefähr 33 Fragen eindeutig identifiziert. Ich habe das Gefühl, dass diese Art der Gefahren vielen nicht bewusst ist.

Dazu kommen die vorher genannten Verknüpfungsangriffe auf Statistiken, die das US-Zensusbüro kürzlich versuchsweise für seinen eigenen Zensus aus dem Jahr 2010 veröffentlicht hat. In diesem Angriff konnten mit Hinzunahme von öffentlichen Daten Einzelpersonen aus den Statistiken des Zensus deanonymisiert werden.

Was heißt das für uns?

Hier muss klar sein: Perfekte Anonymisierung ist unmöglich, wenn man zum Beispiel möchte, dass die Daten sinnvoll bleiben. In der Praxis wird immer nur eine annähernde Anonymisierung möglich sein. Wenn genügend Statistiken über eine Person oder über die gleiche Personengruppe veröffentlicht sind, die verschiedene Eigenschaften einer Person oder auch medizinische Symptome einer Person beleuchten, dann kann man daraus neue Erkenntnisse über die Person gewinnen.

Wo setzen Sie da an? Was ist Ihr Ziel mit Ihrem Projekt und mit Ihrer Forschung?

Das AnoMed-Kompetenzzentrum verfolgt mehrere Ziele. Eines der Hauptziele ist, bessere Anonymisierungstechnologien zu bauen. Die Verfahren, die man vor zehn Jahren benutzt hat und die auch heute noch viel verwenden werden, wie Datenreinigung, K-Anonymität, L-Diversity, funktionieren nicht, wenn die gleichen Daten mehrfach benutzt werden. Und das ist genau das, was beim Gesundheitsdatennutzungsgesetz (GDNG) und beim EHDS passieren soll. Die gleichen Daten können und werden in mehreren Studien genutzt werden. Unsere Techniken zur Anonymisierung setzen nicht an den einzelnen Daten an, sondern daran, die Ergebnisse dieser Studien sicher zu anonymisieren.

Dazu gehört auch, dass wir mit unserem Partner vom Unabhängigen Landeszentrum für Datenschutz Schleswig-Holstein erforschen, welche rechtlichen Implikationen unsere Forschung hat. Also inwiefern ist es denn anonym im Sinne der DSGVO und inwiefern können die Anforderungen, die im GDNG und in anderen Gesetzen formuliert sind, durch das erreicht werden, was wir technisch bauen. Dazu gehören auch klassische IT-Sicherheitsfragestellungen.

Außerdem wollen wir eine Wettbewerbsplattform bauen, um die internationale Forschungsgemeinschaft auf medizinische Fragestellungen aufmerksam zu machen und sie zu motivieren, Lösungen für die Fragestellungen, die wir formuliert haben, zu geben.

Das Projekt ist erst etwa zur Hälfte rum. Was haben Sie bisher erreicht?

Das Projekt geht an vielen Fronten gut voran, aber zur Mitte des Projekts ist der Großteil der Arbeiten noch nicht veröffentlichungsreif. Ich nenne aber gern ein paar Zwischenergebnisse.

Perfekte Anonymisierung ist unmöglich, ohne alle Information zu zerstören. Deswegen sind Anonymitätsgarantien notwendigerweise imperfekt. In den letzten Jahren gab es viele Ansätze Anonymitätsgarantien zu charakterisieren (wie K-Anonymität), die sich als unzureichend herausgestellt haben. Der aktuell in großen Teilen der Forschung als Goldstandard akzeptierte Standard für die Charakterisierung von Anonymisierung heißt Differential Privacy und basiert auf dem Begriff der plausiblen Abstreitbarkeit für einen Datenverarbeitungsalgorithmus. Dieser Begriff ist allerdings sehr unintuitiv. Wir haben signifikante Fortschritte dabei gemacht, Werkzeuge für die Verständlichkeit von Differential Privacy zu entwickeln, indem wir eine mathematische Verknüpfung zu einem der K-Anonymität verwandten Begriff beweisen. Wir haben erste vielversprechende Resultate für Anonymisierung durch Datensynthese (genauer Differentially Private Data Synthesis) erreicht.

Diese und weitere Resultate haben wir auf der Anomed-Seite veröffentlicht.

Wie kompliziert ist es, das Ganze umzusetzen?

Es ist schon noch sehr komplex. Aber es ist heute natürlich schwieriger als es morgen sein wird, und morgen schwieriger als übermorgen. Wir sind noch am Anfang, das wirklich einfach zugänglich zu machen. Da liegt noch viel Arbeit vor uns, aber es passiert auch gerade sehr viel. Es gibt verschiedene Firmen und Initiativen, die versuchen, existierende Lösungen einfacher handhabbar zu machen.

Es gibt viele Diskussionsrunden von Politikern, Fachleuten und Akteuren im Gesundheitswesen zum Thema Datensicherheit. Was denken Sie, wenn die Bedenken der Datenschützer mit dem Verweis auf die Anonymisierung entkräftet werden sollen?

Ich denke, wir sollten uns den Herausforderungen der Digitalisierung des Gesundheitswesens und der weiten Öffnung für medizinische Forschung mit offenen Augen stellen. Die Gesundheitsversorgung und damit verbundene medizinische Forschung hat fraglos höchste Priorität. Gleichzeitig müssen wir die Risiken, denen wir uns mit einer derart durchdringenden Digitalisierung aussetzen, ernst nehmen und aktiv angehen. Wir sollten z.B. so schnell wie möglich alle Löcher stopfen, die sich auftun. Das braucht viel Begleitforschung und Begleitanalyse, das braucht Erforschung neuer Verfahren, das braucht viel IT-Sicherheitsengineering.

Eines der neuen Risiken ist die zentrale Speicherung von Patientendaten aus der medizinischen Versorgung. Aus IT-Sicherheitsperspektive gehen wir ein großes Risiko ein, wenn wir Patientendaten aus der Versorgung zentral speichern und sie nicht mehr nur verteilt irgendwo in der Akte oder auf irgendwelchen Rechnern von Ärzten liegen. Wenn Kriminelle nur wenige zentrale Systeme und Organisationen angreifen müssen, um an alle Patientendaten Deutschlands zu kommen, können sie ihre Aktivitäten auf diese zentralen Systeme und deren Mitarbeiter speziell zuschneiden. Damit kann jeder kleine Fehler, der bei der Aufbewahrung der elektronischen Patientenakte zum Beispiel passiert, verheerend sein. Damit lastet eine große Verantwortung auf die Sicherung dieser zentralen Systeme und deren Organisationen.

Diese Risiken sollten nicht nur den Experten vor Ort bewusst sein. Mündige Bürger:innen in einer Demokratie sollten sich dieser Gefahren ebenfalls bewusst sein, damit sie eine proaktive Sicherung, eine einladende kontinuierliche Begleitanalyse und Begleitforschung fordern. Ich bin mir nicht sicher, ob wir uns als Gesellschaft bewusst sind, welchem Risiko wir da gerade entgegengehen. Ich denke, wir sollten beides tun: Medizinische Studien ermöglichen und uns bewusst machen, wo wir noch nicht perfekt sind, und daran arbeiten.

©Privat

ZUR PERSON

Prof. Esfandiar Mohammadi ist Professor an der Universität zu Lübeck. Er forscht an mathematisch charakterisierbaren Anonymisierungstechniken und an anonymer Kommunikation. Er ist wissenschaftlicher Leiter des BMBF-Kompetenzzentrums AnoMed, dessen Fokus die Forschung an Anonymisierungsfragestellungen für medizinische Anwendungen und insbesondere für maschinelle Lernverhalten ist.

25.08.2024 07:41, Autor: ea, © änd Ärztenachrichtendienst Verlags-AG
Quelle: https://www.aend.de/article/230535