VORAUSSETZUNGEN FÜR EINE PRAXISTAUGLICHE, KI-GESTÜTZTE ANONYMISIERUNG AN GERICHTEN UND VERWALTUNGEN

Andrea Schmidheiny Konic/Bojan Konic

Andrea Schmidheiny Konic, lic. iur., Balo.ai GmbH
Mühleweiher 6, 8606 Greifensee, CH
andrea.schmidheiny@balo.ai, https://www.balo.ai

Bojan Konic, MSc ETH Zürich, Balo.ai GmbH
Mühleweiher 6, 8606 Greifensee, CH
bojan.konic@balo.ai, https://www.balo.ai

Schlagworte: ***Anonymisierung, Künstliche Intelligenz, Gerichte, Verwaltungen***

Abstract: *Das Thema Künstliche Intelligenz ist noch immer stark durch seinen akademischen Ursprung geprägt. Dasselbe gilt auch für Künstliche Intelligenz im Rechtswesen: Was unter klinischen Bedingungen hervorragend funktioniert, besteht die Feuertaufe in der Praxis selten. Eine KI-gestützte Anonymisierungslösung für Gerichte und Verwaltungen drängt sich auf, doch welche Punkte müssen erfüllt sein, damit diese unter realen Bedingungen eine erhebliche Zeitersparnis bei der Anonymisierung von Gerichts- und Verwaltungsentscheiden ermöglicht? Dieser Aufsatz beschreibt die drei wichtigsten Erfolgsfaktoren, damit eine KI-gestützte Anonymisierung tatsächlich funktionieren kann: Zeitersparnis, Anpassbarkeit und Datenschutz. Er fasst die wichtigsten Erkenntnisse aus jahrelanger Entwicklung und repräsentativen Praxistests zusammen.*

1. Einführung in die Anonymisierung

Art. 6 Abs. 1 EMRK postuliert das Öffentlichkeitsprinzip als Grundsatz des Prozessrechtes und erwähnt dabei die Verhandlungsöffentlichkeit sowie die Öffentlichkeit der Urteilsverkündung. Die meisten europäischen Staaten haben diesen Grundsatz in ihren Verfassungen sowie in entsprechenden (Prozess-) Gesetzen weiter konkretisiert und die Publikation von Gerichtsentscheiden vorgesehen. Damit sollen Gerichtsentscheide allen interessierten Personen öffentlich zugänglich gemacht werden. Das Veröffentlichen von Gerichtsentscheiden liegt im öffentlichen Interesse. Dem gegenüber stehen die schutzwürdigen Interessen von Verfahrensbeteiligten sowie weiterer Dritter auf Geheimhaltung. Von wenigen Ausnahmen abgesehen, werden daher zur Veröffentlichung vorgesehene Entscheide in aller Regel in anonymisierter Form veröffentlicht. Um den schutzwürdigen Interessen Verfahrensbeteiligter und Dritter Genüge zu tun und die Lesbarkeit und damit die Verständlichkeit der Entscheide zu gewährleisten, werden Personen, Organisationen und Ortschaften sowie weitere identifizierende Begriffe, wo notwendig, durch Platzhalter ersetzt.

Das Prinzip mag vielerorts dasselbe sein, wirft man jedoch einen genaueren Blick auf anonymisierte Entscheide, wird deutlich, dass sich die konkrete Anonymisierungspraxis von Gericht zu Gericht stark unterscheidet. An Gerichten ohne verschriftlichte Regeln findet man zudem innerhalb desselben Gerichts unterschiedliche Platzhalterformate und Darstellungsformen. Einige Gerichte schwärzen zu anonymisierende Begriffe, wodurch aber meist der Kontext innerhalb des Entscheides im anonymisierten Entscheid nur noch spärlich nachvollzogen werden kann. Faktisch lässt sich eine solcher Entscheid nicht mehr lesen, geschweige denn verstehen. Dazu sei bemerkt, dass mit der alleinigen Veröffentlichung eines Entscheides dem Öffentlichkeitsprinzip noch nicht Genüge getan ist. Die Verständlichkeit des Inhalts muss ebenfalls gewährleistet sein.[1]

[1] BGE 133 I 106, E. 8.3.

2. Bisherige Anonymisierung

Die gängigste Vorgehensweise bei der Anonymisierung von Entscheiden ist die Funktion "Suchen und Ersetzen" in Microsoft Word. Diese Methode hat diverse Nachteile: Werden Namen mit einem weichen Zeilenumbruch getrennt (in Word [Shift] und [Enter]), ist die getrennte Suche nach Vor- und Nachnamen erforderlich. Auch nachträgliche Korrekturen sind nur schwer vorzunehmen, da das Ersetzen der Begriffe sukzessive und endgültig ist. Durchschnittlich werden mit dieser Methode pro Entscheid zirka 46 Minuten für die Anonymisierung aufgewendet.[2]

3. Künstliche Intelligenz und Anonymisierung

Eine Disziplin der Künstlichen Intelligenz befasst sich insbesondere mit dem Verständnis von Sprache und Texten: Das «Natural Language Processing» – auch NLP genannt. NLP wiederum lässt sich in weitere Unterthemen aufgliedern. Für die Anonymisierung besonders relevant ist das Erkennen von Entitäten – oder «Named Entity Recognition». Damit ist es möglich, Entitäten zu erkennen, wie z.B. Personennamen in einem Text (oder Entscheid), ohne eine Liste von bekannten Namen hinzuziehen zu müssen. Die Technologie ermöglicht es, dass das System durch wiederholtes Trainieren ein implizites Verständnis davon gewinnt, was z.B. den Namen einer Person, einer Organisation oder einer Ortschaft ausmacht. Das «Training» besteht daraus, dem System mehrere tausende Beispiele zu zeigen. Dadurch wird ein darunterliegendes statistisches Modell angepasst und optimiert. Analog dazu können im Anschluss auch weitere Kategorien von Begriffen trainiert und erkannt werden.

Dank der Named Entity Recognition ist es möglich, Personen, Organisationen und Ortschaften automatisch zu erkennen. Der überwiegende Teil der Fleissarbeit wird damit von der Künstlichen Intelligenz geleistet. Der anonymisierenden Person wird damit innert kurzer Zeit ein Überblick über die zu anonymisierenden Elemente verschafft.

Eine vollautomatische Anonymisierung dagegen – also eine Anonymisierung von Gerichtsentscheiden ohne Kontrolle und allfällige Ergänzung durch eine Fachperson – ist nicht möglich und wird wohl auch in den nächsten Jahrzehnten unrealistisch bleiben. Natürlich ist es aus technischer Perspektive möglich, eine vollautomatische Anonymisierung vorzunehmen. Entweder wird jedoch die Verständlichkeit des Textes beträchtlich eingeschränkt und somit das Öffentlichkeitsprinzip verletzt oder es besteht das Risiko, dass identifizierende Elemente nicht erkannt werden, obwohl die Kombination von mehreren nicht anonymisierten Elementen zu einer Identifikation führt. Letztlich bleibt es die Aufgabe der Fachperson, die in Ziffer 1 beschriebene Interessenabwägung vorzunehmen. Sie muss abwägen, wo das öffentliche Interesse an Information überwiegt und welche Information im konkreten Entscheid unter Berücksichtigung sämtlicher weiterer Angaben identifizierend sein kann. Da Angaben in jedem Sachverhalt auf andere Weise identifizierend sein können, bleibt eine qualitativ ausreichende vollautomatische Anonymisierung ein Luftschloss.

Nebenbei sei erwähnt, dass allfällige sich aus einer vollautomatischen Anonymisierung ergebende Haftungsfragen zunächst noch zu klären wären. Eine erste Auseinandersetzung mit der Problematik findet sich in TANIA MUNZ, Staatshaftung für mangelhafte Anonymisierung von publizierten Gerichtsurteilen, in: «Justice – Justiz – Giustizia» 2022/1.

[2] DANIEL HÜRLIMANN, Was kostet die Anonymisierung von Urteilen? in: Daniel Hürlimann/Daniel Kettiger (Hrsg.), Anonymisierung von Urteilen, Helbing Lichtenhahn Verlag, 2021, Rz. 3.

4. Theorie und Praxis

Man möchte in Anbetracht der technischen Möglichkeiten denken, es sei ein Leichtes, die «Named Entity Recognition» auf Gerichtsentscheide anzuwenden. Es hat sich jedoch bei der Entwicklung des A-Tools rasch gezeigt, dass sich die Sprachfacetten von Gerichtsentscheiden sehr stark von denjenigen Texten unterscheiden, die üblicherweise für das Training von Named Entity Recognition-Modellen verwendet wurden.

Dies liegt mitunter auch an der Verfügbarkeit der Input-Texte: Anonymisierte Gerichtsentscheide sind nicht in demselben Umfang vorhanden wie Beiträge in offenen Wissenssammlungen oder Foren. Es überrascht daher auch nicht, dass vorhandene NER-Modelle auf Wikipedia-Einträge oder Kommentaren aus Online-Foren wie «Reddit» trainiert sind. Dies hängt auch damit zusammen, dass das Annotieren dieser Texte – also das manuelle Markieren der Entitäten in Texten – sehr aufwendig ist. Häufig werden bereits annotierte und offen verfügbare Texte verwendet, um Basismodelle zu trainieren, damit verschiedene NLP-Systeme auch besser miteinander verglichen werden können.

Die juristische Sprache in Gerichtsentscheiden unterscheidet sich stark von den oft umgangssprachlichen Kommentaren in Onlineforen. Dies ist in den trainierten Modellen nicht nur bei der Qualität der erkannten Entitäten spürbar. Bereits die dazu notwendige Aufteilung von Sätzen in einzelne Einheiten («Tokens»), die im häufigsten Fall aus den einzelnen Worten bestehen, ist grundsätzlich anders.

Ein Beispiel:

In einem Gerichtsurteil wurde in einer Klammer eine Adresse erwähnt: «…Bahnhofstrasse 8)». Im ursprünglichen Modell wurde «8)» – also die Zahl «8» und die schliessende Klammer – zusammen als eigener Token erkannt. Der Grund dafür war, dass in Online-Kommentaren oft Emojis verwendet werden und «8)» für ein Smiley mit Sonnenbrille steht. (😎). Es darf mit grosser Wahrscheinlichkeit angenommen werden, dass Emojis in Gerichtsurteilen nur sehr selten Anwendung finden. Entsprechend wurde für das A-Tool ein eigenes, für Gerichtsurteile ausgelegtes Modul («Tokenizer») geschrieben, welches diese Trennung der Einheiten korrekt vornimmt.

Ein weiterer Unterschied sind die zahlreichen Abkürzungen, die in Entscheiden verwendet werden. Während es in der englischen Sprache unüblich ist, wird in der deutschen Schriftsprache eine Abkürzung oft mit einem Punkt («.») beendet. Ohne weitere Anpassung erkennen einige NLP-Systeme eine Abkürzung meist nicht korrekt und halten es für das Ende des jeweiligen Satzes. Dies führt dazu, dass Begriffe nach der Abkürzung nicht mehr korrekt in das Training einbezogen werden und das Modell somit «falsch trainiert» wird. Es musste somit ein Weg gefunden werden, wie Abkürzungen als solche korrekt und nicht als Ende eines Satzes erkannt werden.

Letztlich ist es essenziell, dass NER-Modell nicht nur generell auf deutschsprachige Gerichtsentscheide im Allgemeinen sondern auch auf spezifische Gerichtsurteile des jeweiligen Gerichtes trainiert und getestet werden. Nur so kann eine genügend hohe Qualität der KI gewährleistet werden. Balo.ai trainiert ein spezifisches NER-Modell für jeden einzelnen Kunden, um die bestmögliche Qualität zu gewährleisten.

5. Voraussetzungen

Die Erfahrung mit dem A-Tool hat gezeigt, dass insbesondere drei Voraussetzungen gegeben sein müssen, damit eine KI-gestützte Anonymisierungslösung in der Praxis funktioniert. Die einzelnen Punkte werden in der Folge detailliert ausgeführt:

- Zeitersparnis bei der Anonymisierung
- Anpassbarkeit der Anwendung auf das jeweilige Gericht
- Sicherstellung des Datenschutzes

6. Zeitersparnis

Es ist zwar naheliegend, dass eine Anwendung bei der Anonymisierung Zeit sparen soll, jedoch keineswegs selbstverständlich: Wie Umfragen ergeben haben, waren bisher mehrheitlich Anwendungen im Einsatz, die keine spürbare Zeitersparnis brachten.[3]

Mit der Verwendung von NLP können grundsätzlich sämtliche Entitäten erkannt und somit ein wesentlicher Anteil an Aufwand eingespart werden. Allerdings liegen jeder Künstlichen Intelligenz statistische Methoden zugrunde, welche vollumfänglich auf den Trainingsdaten basieren. Ein NER-Modell kann somit immer nur folgende Aussagen machen: «Aufgrund der Daten, die ich bisher gesehen habe, ist Wort X an Stelle Y mit hoher Wahrscheinlichkeit eine Entität des Typs Z».

Wir müssen folglich annehmen, dass es auch Gerichtsentscheide geben kann, zu welchen auch ein sehr aufwendig trainiertes NER-Modell fehlerhafte Aussagen bezüglich der Entitäten macht. Nebenbei ist dies ein weiterer Grund, wieso eine qualitativ zufriedenstellende vollautomatische Anonymisierung unrealistisch ist. Es ist daher von grösster Bedeutung, dass die Benutzer der Anwendung Ungenauigkeiten oder Fehler beheben können, ohne dabei viel Zeit einzubüssen. Eine bedienerfreundliche und intuitive Benutzeroberfläche ist dabei essenziell.

Für die Effizienzsteigerung ist zudem die optimale Unterstützung der vorhandenen Arbeitsprozesse entscheidend. Ein Entscheid, der in Microsoft Word geschrieben wurde, sollte idealerweise auch direkt in Word anonymisiert werden können.

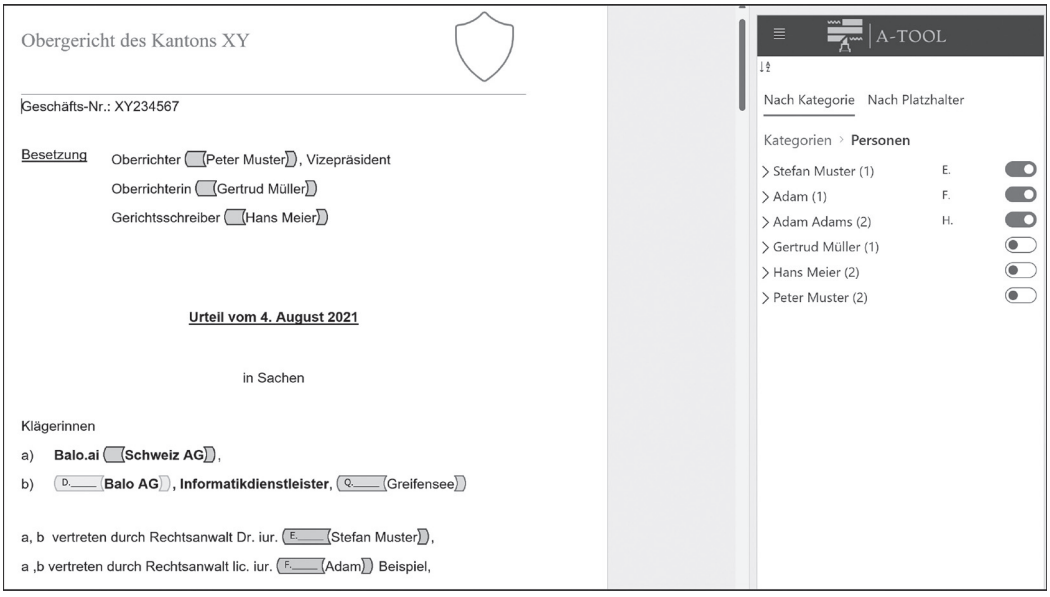

Abbildung 1: Screenshot aus Word eines analysierten Entscheides in A-Tool

Durch eine übersichtliche Auflistung der Entitäten entfällt ein erstes Durchlesen des Entscheides, Korrekturen im Rahmen der Interessenabwägung werden in wenigen Schritten vorgenommen, womit die aufwendigen Fleissarbeiten durch die Anwendung erledigt werden. Idealerweise finden Texterstellung und Anonymisie-

[3] Daniel Hürlimann/Daniel Kettiger, Zugänglichkeit zu Urteilen kantonaler Gerichte: Ergebnisse einer Befragung, in: «Justice – Justiz – Giustizia» 2018/2, Rz. 16.

rung in derselben Textbearbeitungssoftware statt, z.B. in Microsoft Word. Dadurch kann bis zu 70 % des Anonymisierungsaufwandes eingespart werden.

7. Anpassbarkeit

Wie unter Ziffer 4 bereits erwähnt, ist es sinnvoll, die KI ganz konkret auf Entscheide des spezifischen Gerichtes zu trainieren, um die bestmögliche automatische Erkennung der Entitäten zu gewährleisten. Zu berücksichtigen gilt darüber hinaus, dass sich die Anonymisierungspraxis, z.B. die Verwendung von bestimmten Platzhaltern für zu anonymisierende Begriffe, ebenfalls massgeblich unterscheiden kann.

So verwenden beispielsweise einige schweizerische Ober- bzw. Kantonsgerichte aufsteigende Buchstaben mit Punkt und Unterstrichen als Platzhalter (z. Bsp. «A._____») während österreichische Gerichte oft aufsteigende Nummern mit Sternen davor und danach einsetzen. (z. Bsp. «***1***»). Ob und wie Daten (wie Geburtsdaten), E-Mail-Adressen, KFZ-Nummern, IBAN-Nummern etc. ersetzt werden, wird zuweilen von Gericht zu Gericht unterschiedlich gehandhabt. Zuletzt müssen Struktur und Darstellung der Entscheide berücksichtigt werden. Gerichte gestalten zudem Deckblätter, Rubren, Dispositive und Unterschriftsblöcke unterschiedlich. Eine optimale Unterstützung bei der Anonymisierung ist daher nur dann gegeben, wenn dies kundenspezifisch berücksichtigt wird.

Eine konfigurative Anpassung und Erweiterung der erkannten Entitäten und deren Platzhalter-Logik muss ohne Programmierung in der Anwendung möglich sein. Zudem müssen relevante Textstellen wie Rubrum, Dispositiv oder Unterschriftenblock korrekt erkannt und gemäss der geltenden Anonymisierungspraxis des jeweiligen Kunden berücksichtigt werden.

8. Datenschutz

Es liegt auf der Hand, dass die sensiblen Informationen in nicht anonymisierten Gerichtsentscheiden bei der Verarbeitung durch eine Künstliche Intelligenz besonders geschützt werden müssen. Wir empfehlen daher, diese Verarbeitung ausschliesslich in der IT-Umgebung des Gerichtes stattfinden zu lassen und die notwendige Infrastruktur der Anwendung daher auch intern («on premises») zu bewirtschaften. Auf diese Weise verlassen die Daten niemals die gesicherte Gerichtsumgebung.

Eine Verarbeitung in einer (Private- oder Public-) Cloud-Umgebung wäre rein technisch natürlich möglich. Es gibt jedoch in Bezug auf den Datenschutz eine Vielzahl offener Fragen, die in Österreich, Deutschland und der Schweiz aktuell auf mehreren Ebenen diskutiert werden.

Es hat sich in der Praxis bewährt, eine solche Anwendung innerhalb der Gerichtsumgebung zu installieren. Durch eine verschlüsselte Verbindung und die Einbettung in der Gerichtsumgebung ist der Datenschutz jederzeit gewährleistet.

9. Referenz: Gerichte und Verwaltung des Kantons Aargau

Das A-Tool ist für Entscheide in deutscher Sprache als Add-In für Word verfügbar. Demnächst erscheint A-Tool für PDF, sowie die Unterstützung beider Produkte für Entscheide in französischer und italienischer Sprache.

Die Gerichte und die Verwaltung des Kantons Aargau setzen das A-Tool von Balo.ai seit Januar 2022 erfolgreich ein. Es wurden bisher über 100 Personen in der Verwendung des Tools geschult. Sämtliche Sachentscheide werden seit der Einführung innerhalb von 10 Tagen nach Entscheidfällung anonymisiert und im Internet veröffentlicht.