

Ohne Menschen können die Maschinen nicht lernen

Der Einsatz von KI im Bundesarchiv

geschrieben von Andrea Hänger

Die Frage nach dem Einsatz von Künstlicher Intelligenz (KI) im Archiv stellt sich, seitdem Archive in immer größer werdendem Umfang ihre analogen Akten, Bilder oder Filme digitalisieren. Damit entstehen erstmals relevante Datenmengen, die mit Methoden der Künstlichen Intelligenz besser und einfacher erschlossen und ausgewertet werden können. Bisher wurde mit der Digitalisierung die digitale Nachbildung der analogen Welt vollzogen. Anstelle der Originale im Lesesaal können nun die Abbilder der Akten am eigenen Rechner durchgeblättert werden. Der Gewinn liegt im Einsparen der Reisezeit und damit der orts- und zeitungebundenen Benutzung sowie der permanenten Verfügbarkeit. Doch was zunächst noch Luxus und neuer Service war, stieß auch schnell an seine Grenzen und ließ neue Wünsche zum Beispiel nach durchsuchbaren Volltexten entstehen. Diese fehlen, weil das in anderen Bereichen durchaus erfolgreiche Verfahren der optischen Zeichenerkennung (OCR) bei typischem Archivgut mit wechselnden Schriftarten auf jeder Seite sowie einer Mischung von hand- und maschinengeschriebenen Texten nur ungenügende Ergebnisse erzielt. Erst mit dem Einsatz künstlicher neuronaler Netze kann die Erkennung solcher Texte erheblich verbessert werden.

Das Bundesarchiv startete sein erstes Projekt zur Handschriftenerkennung mithilfe von KI mit den besonders schwer lesbaren Akten des Reichskolonialamtes. Verschiedene Handschriften in Kurrentschrift mit zahlreichen Marginalien auf jeder Seite prägen die Überlieferung. Der Einsatz von KI soll hier nicht nur eine verbesserte Texterkennung ermöglichen und damit die Volltextrecherche unterstützen, sondern kann auch fehlende Methoden- und Sprachkenntnisse der Nutzenden ersetzen. Das Bundesarchiv hat schon seit einigen Jahren sämtliche Unterlagen zur deutschen Kolonialherrschaft digitalisiert und frei online zugänglich gemacht. Der damit angestrebte Zugang für alle, gerade auch für die Menschen in den ehemaligen Kolonien, die etwas über die Geschichte ihrer Heimat erfahren möchten, ist damit zwar technisch möglich. Die doppelte Hürde der fremden Sprache und Schrift, die auch für einen Großteil der die deutsche Sprache beherrschenden Nutzenden unlesbar ist, begrenzt die tatsächlichen Auswertungsmöglichkeiten jedoch extrem. Mit dem KI-Projekt sind nun rund 10.000 Akten des Reichskolonialamtes vollständig durchsuchbar und maschinenlesbar geworden. Damit ist es jetzt auch möglich, sie über im Internet verfügbare Programme zu übersetzen.

In einem Modelltraining haben mit der Schrift und der Überlieferung vertraute Archivarinnen und Archivare des Bundesarchivs immer wieder von der KI gelesene Texte korrigiert und so Erkennungsraten von bis zu 90 Prozent erreicht. Die KI-gestützte Recherchemöglichkeit wird in Kürze in den Lesesälen des Bundesarchivs Interessierten zur Verfügung stehen. Die Beschränkung auf den Lesesaal liegt nicht nur an den hohen Kosten für die im Moment noch nach Zugriffszahlen kalkulierten Lizenzen. Sie hat ihren Grund auch in der noch fehlenden

Kontextualisierungsmöglichkeit in einem rein virtuellen Raum. Weil die Sprache der Quellen durch und durch rassistisch ist, benötigt sie einen erklärenden Rahmen, der sich bei der Benutzung einer Originalakte aus dem Reichskolonialamt allein schon durch die äußere Anmutung, den Geruch, die Schrift, die Haptik und den konkreten Raum des Archivs von selbst herstellt.

Nach den positiven Erfahrungen des ersten Projektes hat das Bundesarchiv, gefördert durch Sondermittel des Bundes, ein eigenes KI-Programm aufgelegt, in dessen Rahmen eine allgemeine KI-Strategie erarbeitet wird und weitere Projekte durchgeführt werden. Der Schwerpunkt liegt hier auf der KI-basierten Erschließung von Karteien, um personenbezogene Anfragen schneller bearbeiten zu können bzw. überall dort, wo die rechtlichen Möglichkeiten bestehen, interessierte Bürgerinnen oder Forscher selbst online recherchieren zu lassen.

Welche Möglichkeiten solche Projekte bieten, zeigt ein besonders spannendes Projekt im ungarischen Nationalarchiv. Ein wichtiges, bisher kaum über Quellen zugängliches Thema in der ungarischen Gesellschaft ist das Schicksal von mehr als 600.000 Ungarinnen und Ungarn in sowjetischer Kriegsgefangenschaft im Zweiten Weltkrieg. Ihr Schicksal wurde über Jahrzehnte verdrängt und verschwiegen und erst 2015 mit dem Gulag-Gedenktag wieder in das Bewusstsein der Öffentlichkeit gebracht. Das ungarische Nationalarchiv ließ die Karteikarten zu den Kriegsgefangenen in russischen Archiven digitalisieren, KI-gestützt auslesen und in eine Datenbank übertragen. Die Herausforderung bestand darin, dass bei der Aufnahme in die Lager sowjetische Militärschreiber die ungarischen Namen, so wie sie sie verstanden, in kyrillischer Schrift auf die Karteikarten schrieben. Mittels der für das Projekt eingesetzten KI-Software gelang es, die Namen zu lesen, in lateinische Schrift umzuwandeln und weitestgehend zu plausibilisieren. Neben der Texterkennung lassen sich KI-basierte Methoden auch für die Erschließung weiterer Archivgutarten einsetzen. Relativ weit entwickelt ist die Bilderschließung mittels KI-gestützter Gesichtserkennung, weitere Pilotprojekte laufen zur Spracherkennung von Ton- und Filmaufnahmen. Neben diesen die Erschließung unterstützenden Projekten gibt es aber auch weitere vielversprechende Möglichkeiten. Das Schweizer Bundesarchiv zum Beispiel erreicht durch eine KI-basierte Anonymisierung von Personennamen eine weitgehende Öffnung seiner Datenbank für die Öffentlichkeit.

Selbstverständlich sind für einen breiteren Einsatz von KI in den Archiven noch Herausforderungen zu meistern. An erster Stelle stehen hier die benötigten Ressourcen finanzieller, aber auch infrastruktureller Art. Nach den Erfahrungen des Kolonialprojektes würde das Bundesarchiv neun Jahre benötigen, um für alle bisher online stehenden digitalisierten Akten Volltexte zu erzeugen und durchsuchbar zu machen. Selbstverständlich lassen sich die Prozesse parallel durchführen, aber die dafür erforderlichen Rechnerkapazitäten sind gewaltig.

Hinzu kommen die hohen Kosten für den Betrieb der Software und das bisher ungelöste Problem, dass KI-Softwarelösungen ausschließlich in der Cloud angeboten werden. Nur auf diese Weise können die Hersteller die extrem kurzen Innovationszyklen wirtschaftlich abbilden. Für den internen Betrieb maßgeschneiderte Lösungen verursachen hohe Kosten für den Auftraggeber, die aber insbesondere wegen der Verarbeitung personenbezogener Daten unvermeidlich sind.

Nichtsdestotrotz überwiegen die Chancen bei Weitem, und viele Ideen warten noch auf ihre Umsetzung. Mit KI können Aufgaben erledigt werden, die sonst gar nicht angegangen würden. Nutzungshindernisse werden abgebaut, und das Training der Software sorgt in den Archiven für einen wichtigen Innovationsschub. Zugleich aktualisiert und wertschätzt es das Fachwissen der Archivarinnen und Archivare, die unverzichtbare Partner der KI-Projekte bleiben. Ohne Menschen können die Maschinen nicht lernen.

Dieser Text ist zuerst erschienen in Politik & Kultur [04/2023](#).