



Digitale Zeitzeugenschaft – Wenn Algorithmen das digitale Gedächtnis übernehmen

Erfahrungen mit künstlicher Intelligenz im Archiv „Deutsches Gedächtnis“

von Almut Leh

„Die Antwort ist 42“

Mindestens die Älteren unter Ihnen werden sich vielleicht an den Roman „Per Anhalter durch die Galaxis“ erinnern, ein Roman mit zeitweiligem Kultstatus, den Douglas Adams 1979 veröffentlichte. Die Jüngeren haben vielleicht 2005 die Verfilmung gesehen. Eine wunderbar skurrile Geschichte, in der es unter anderem einen Super-Computer mit dem schönen Namen „Deep Thought“ gibt, der die

Antwort auf die „große Frage nach dem Leben, dem Universum und allem“ errechnen soll. Er braucht dafür 7,5 Millionen Jahre und eröffnet dann der atemlos lauschenden Menge: „Die Antwort auf die große Frage nach dem Leben, dem Universum und allem lautet ... 42!“. Die Enttäuschung der Zuhörerschaft ist riesengroß, doch „Deep Thought“ erklärt ungerührt, dass die Wesen, die ihm diese Frage gestellt hätten, nie wirklich wussten, was eigent-

lich ihre Frage war. Das sei aber kein Problem. Wenn man ihm ausreichend Zeit gebe, könne er mit einer neuen Version seiner selbst auch die korrekte Frage errechnen, deren Antwort 42 lautet.

Warum erzähle ich das? Weil eine Antwort immer nur so gut ist, wie die zugehörige Frage. Es kommt nicht auf die Antwort an, alles hängt von der Frage ab. Das trifft für jede Erkenntnis zu, aber vielleicht noch einmal mehr, wenn man die Antwort einer Maschine überlässt, selbst wenn diese über künstliche Intelligenz verfügt. Gerade dann muss die Frage exakt und maschinengerecht formuliert werden.

Auch an ein Archiv werden Fragen gestellt. Ist das Archiv deshalb eine Maschine, so ein „Deep Thought“? Immerhin, und wie wir sehen werden, kommt es auch beim Archiv darauf an, die richtigen Fragen zu stellen. Das Archiv ist ein Ort, eine Institution, an dem Wissensbestände aufbewahrt werden mit dem Zweck, dieses Wissen dann bereitzustellen, wenn es zur Beantwortung aufkommender Fragen beitragen kann. Es geht also um die Verarbeitung von Informationen, was in der Informatik die Betrachtung als Maschine nahelegen würde.

Lange Zeit war das Image von Archiven nicht das Beste. Die Vorstellung war geprägt von staubigen, dunklen Kellergelassen, vollgestopft mit Dokumenten in mehr oder weniger guter Ordnung, zwischen denen sich verschrobene Menschen, Archivare, bewegen, die den Kontakt zum wirklichen Leben meiden. Das hat sich mit dem „Memory boom“ seit den 1990er Jahren deutlich verändert. Zusammen mit Bibliotheken und Museen werden Archive inzwischen als Gedächtnisinstitutionen angesprochen. Institutionen, die als ideale Gesamtheit das kulturelle Gedächtnis einer Gemeinschaft repräsentieren. Auch die Digitalisierung hat zu diesem Bedeutungswandel beigetragen, weil die digitale Repräsentation der Originaldokumente deren Verbreitung und damit auch Sichtbarkeit erheblich befördert hat.

Als digitale Daten sind die Dokumente zugleich maschinenlesbar geworden, was erhebliche Auswirkungen auf die Organisation von Arbeitsprozessen im Archiv hat. Gesellschaftlich brisant ist dieser Umstand, weil das kollektive Erinnern gleichsam unter der Hand – will sagen: weitgehend unreflektiert – an Algorithmen delegiert wird, mit anderen Worten: zum Gegenstand von Rechenvorschriften gemacht wird.

Ich möchte im Folgenden den Blick darauf lenken, wie sich im Prozess der Wandlung von Dokumenten zu Daten das Arbeitsergebnis verändert und welche Folgen dies für die Generierung von Wissen hat. Dafür werde ich zunächst das Archiv „Deutsches Gedächtnis“ als Forschungsdatenrepositorium und digitales Gedächtnis vorstellen und daran anschließend Projekte zur Sprachtechnologie und zum Wissensmanagement, die wir gemeinsam mit technologischen Partnern in den letzten Jahren durchgeführt haben. Vor diesem Hintergrund werde ich die Auswirkungen der Digitalisierung auf die biographische Forschung und die Archivierung subjektiver Erinnerungszeugnisse problematisie-

ren und mit Überlegungen zu Perspektiven einer digitalen Hermeneutik schließen.

Das Archiv „Deutsches Gedächtnis“ als Forschungsdatenrepositorium und digitales Gedächtnis

Das Archiv „Deutsches Gedächtnis“ ist eine Einrichtung der FernUniversität in Hagen und gehört zum Institut für Geschichte und Biographie, ein Forschungsinstitut, das im Bereich Oral History und Biographieforschung ausgewiesen ist.¹ Um etwas über die jüngere Vergangenheit zu erfahren, befragen wir Menschen zu ihren Lebensgeschichten und Erfahrungen und analysieren dann diese Interviews als historische Quellen. Seit den frühen 1980er Jahren führt das Institut solche Forschungsprojekte durch, und alle diese Interviews werden archiviert. Insofern ist das Archiv „Deutsches Gedächtnis“ ein Forschungsdatenrepositorium, das Forschungsdaten zur Nachnutzung anbietet. Dabei sind Oral History-Interviews äußerst wertvolle Forschungsdaten. Wertvoll zum einen ganz monetär, weil die Herstellung solcher Interviews teuer ist. Wertvoll aber auch im ideellen Sinn, weil die Interviews nicht wiederholbar sind – die Zeitzeugen sind inzwischen verstorben oder würden sich heute vielleicht anders erinnern. Und schließlich sind sie wertvoll, weil sie für Sekundäranalysen besonders ergiebig sind. Als lebensgeschichtliche, narrative Interviews behandeln sie die gesamte Biographie der Interviewten, nicht nur ein bestimmtes Thema. Das Ergebnis sind mehrstündige Ton- oder Videoaufzeichnungen, die eine Fülle von Informationen bieten, die mit der Primärforschung nicht annähernd ausgeschöpft wurden und die für viele weitere Forschungsfragen mit Gewinn ausgewertet werden können.²

Aktuell umfasst das Archiv 3.085 Interviews aus über einhundert Forschungsprojekten. Ergänzt wird die Interviewsammlung durch subjektive Erinnerungszeugnisse wie Autobiographien, Tagebücher und Briefsammlungen. Das Archiv „Deutsches Gedächtnis“ ist von daher eine Gedächtnisinstitution in zweifacher Hinsicht: einmal in der Bewahrung und Vermittlung von Wissensbeständen, die in ihrer Gesamtheit das kulturelle Gedächtnis einer Gemeinschaft bilden. Zusätzlich aber auch in einem unmittelbaren Sinn, insofern die hier archivierten Wissensbestände selbst Erinnerungen, also Gedächtnisinhalte, sind. Im Archiv „Deutsches Gedächtnis“ kreuzen sich gewissermaßen das kulturelle, das kommunikative und das individuelle Gedächtnis.

So sinnvoll es ist, diese Interviews als Forschungsdaten wie auch als kulturelle Gedächtnisinhalte zu bewahren, so herausfordernd ist diese Aufgabe auch. Das liegt nicht zu-

1 Vgl. Almut Leh, Vierzig Jahre Oral History in Deutschland. Beitrag zu einer Gegenwartsdiagnose von Zeitzeugenarchiven am Beispiel des Archivs „Deutsches Gedächtnis“, in: Westfälische Forschungen. Zeitschrift des LWL-Instituts für westfälische Regionalgeschichte, 65 (2015), S. 255–268.

2 Linde Apel/Almut Leh/Cord Pagenstecher, Oral History im digitalen Wandel. Interviews als Forschungsdaten, in Linde Apel (Hrsg.), Erinnern, erzählen, Geschichte schreiben. Oral History im 21. Jahrhundert (2020, in print).

letzten daran, dass die Interviews auf audiovisuellen Medien aufgezeichnet sind.

Archiviert werden Audio- und Videoaufzeichnungen, außerdem Transkriptionen, Protokolle, rechtliche Erklärungen zur Nutzung und Fotos. Seit etwa 20 Jahren wissen wir, dass analoge Tonaufzeichnungen brutal altern und am Ende unbrauchbar werden. Die Lösung: umfassende Digitalisierungsmaßnahmen. Inzwischen ist dieser Prozess weitgehend abgeschlossen, und was neu produziert wird, ist natürlich per se digital. Anders als ein Dokumentenarchiv ist das „Deutsche Gedächtnis“ somit tatsächlich ein sehr weitgehend digitales Gedächtnis. Dadurch ist es nicht nur für Geistes- und Kulturwissenschaftler:innen interessant, sondern auch für Informatiker:innen, für die unsere qualitativen, audiovisuellen Interviews unstrukturierte, multimodale Daten sind, und das in einer Menge, die sich durchaus zum Training neuronaler Netze eignet. Über die Digitalisierung der Bestände hat künstliche Intelligenz Eingang in das nunmehr digitale „Deutsche Gedächtnis“ gefunden. Zwei Projekte, die daraus entstanden sind, will ich näher vorstellen.

Technologische Projekte zur Sprachtechnologie und zum Wissensmanagement

In den letzten fünf Jahren haben wir zusammen mit dem Fraunhofer Institut für intelligente Analysensysteme ein sprachtechnologisches Forschungsprojekt durchgeführt mit dem Ziel, die automatische Spracherkennung zu verbessern.³ Wieso verbessern? Funktioniert doch, mag mancher denken. Doch was im Zwiegespräch mit Siri, Alexa und Co gut läuft, lässt sich nicht auf Oral History-Interviews übertragen. Hier haben wir es mit undeutlicher, oft dialektaler Aussprache zu tun, mit überlappenden Sprecherwechseln, mit suboptimalen Audioaufzeichnungen und den Folgen zu spät digitalisierter Magnetbänder.

Zu Beginn des Projektes lag die Fehlerquote der automatischen Spracherkennung je nach Qualität der Aufzeichnung bei bis zu 55 Prozent. Nach fünf Jahren Forschungsarbeit lag die Wortfehlerquote im Mittel bei 25 Prozent. Das hört sich immer noch viel an, ist in der Praxis aber schon sehr brauchbar, weil vieles, was als Fehler qualifiziert wird, das Textverständnis nicht beeinträchtigt.

Tatsächlich ist die automatische Spracherkennung für die Archivierung von Zeitzeugeninterviews ein großer Gewinn, und zwar vor allem bei der Recherche nach Interviews, die zur Bearbeitung bestimmter Forschungsfragen relevant sein können. Diese Recherche ist eine der schwierigsten Aufgaben und zugleich natürlich unabdingbar, wenn denn Archivierung sinnvoll sein soll. Archivieren heißt schließlich nicht sammeln und vor jedem Zugriff schützen, sondern nutzbar machen und bereitstellen. Aber wie findet man die zu einer Forschungsfrage passenden Quellen, in unserem Falle Interviews?

Ein wichtiges Rechercheinstrument ist die Volltextsuche nach bestimmten Begriffen, was nach bisherigem technischen Stand in den Transkripten geschieht. Die automati-

sche Spracherkennung ermöglicht nun nicht nur die automatische Herstellung von Transkripten, sondern auch die Recherche unmittelbar im Audiosignal, eine Technologie, die analog zum Textmining als Audio Mining bezeichnet wird. Das Audio Mining-System des Fraunhofer Instituts bietet eine Medienansicht, in der die Medien mit Untertiteln abgespielt werden. Man kann parallel zum Hören bzw. Sehen im Transkript lesen oder auch das Transkript downloaden. Man kann in Freitext nach Begriffen oder Wortfolgen suchen, die in einer Trefferliste im Kontext angezeigt werden. Bei Auswahl eines Treffers kann man direkt an die betreffende Stelle im Audio oder Video springen und sich die Sequenz anhören. Außerdem werden für einzelne Interviews automatisch Keywords generiert, die eine erste Orientierung über den Inhalt ermöglichen.

Für die Recherche nach Interviews ist diese Technologie sehr hilfreich, in mancher Hinsicht vielleicht sogar ein Durchbruch. Auch in einem recht gut geführten Archiv wie dem „Deutschen Gedächtnis“ ist etwa ein Drittel der Interviews nicht transkribiert. Mit Hilfe des Audio Mining können diese nicht-transkribierten Interviews in die Suche einbezogen werden. Auch in der synchronen Darstellung von Text und Audio- bzw. Videoaufzeichnung sehe ich einen großen Gewinn. Bisher wurden Interviews in der Forschungspraxis meist auf einen Text reduziert. Jetzt können auch Prosodie, Mimik und Gestik analysiert werden – auch wenn dies unser Auswertungsinstrumentarium aktuell noch überfordert. Aber auch dabei kann vermutlich künstliche Intelligenz schon jetzt oder in naher Zukunft weiterhelfen, indem Interviews nicht nur nach Worten, sondern auch nach Tonmodulationen, Sprechgeschwindigkeit, Pausenlängen, Gesten und anderen nicht-sprachlichen Elementen durchsucht und analysiert werden. Auch werden wir mit technischer Unterstützung eine deutlich größere Zahl von Interviews in eine Untersuchung einbeziehen können.

In einem weiteren Projekt geht es um die Verfügbarmachung der Interviews. Mit der fortschreitenden Digitalisierung wurde die Forderung der Nutzer:innen nach Übermittlung von Transkripten, Audio- und Videodateien via Internet immer lauter und drängender. Technisch ist das in der Tat kein Problem. Und was liegt näher, als Oral History-Interviews in einem Online-Archiv zugänglich zu machen? Die rechtlichen Aspekte bezogen auf Daten- und Persönlichkeitsschutz der Interviewten lasse ich einmal außen vor.

Gemeinsam mit Kolleg:innen an der Freien Universität Berlin und anderen Partnern bauen wir derzeit ein solches Online-Angebot auf. Das Projekt heißt Oral History.digital. Das Ziel ist die Bereitstellung einer digitalen Informationsinfrastruktur, ein zentraler Zugangspunkt, der Forschenden die Möglichkeit gibt, online in einer Vielzahl von Interview-

³ Joachim Köhler/Michael Gref/Almut Leh, Systeme der Spracherkennung im Kontext der Oral History, in: BIOS. Zeitschrift für Biographieforschung, Oral History und Lebensverlaufsanalysen, 30 (2017), S. 44–59, <https://doi.org/10.3224/bios.v30i1-2.05> [Stand: 19.08.2021, gilt ebenfalls für alle weiteren Hinweise auf Internetseiten].

beständen zu recherchieren und die ausgewählten Interviews direkt anzuhören, zu annotieren und die Ergebnisse downloaden zu können.⁴

Schon jetzt, vier Monate nach Beginn des Projektes Oral History.digital, zeichnet sich ab, dass das Projekt ein großer Erfolg werden kann. Das Interesse von Institutionen, die Oral History-Interviews haben und in Oral History.digital bereitstellen wollen, ist riesig und übertrifft unsere Erwartungen. Und was die Forschung betrifft, wird Oral History.digital die Nutzung von Oral History-Interviews nochmals attraktiver machen. Gerade in den letzten Jahren ist ein verstärktes Interesse an der Quelle unübersehbar geworden. Dabei wird die Sekundärauswertung von Interviews immer wichtiger. Wenn man untersuchen will, wie Menschen in Ost und West vor der Wiedervereinigung auf das je andere Deutschland geblickt haben, braucht man Interviews, die vor 1989 geführt wurden. Wenn man Nationalsozialismus und Zweiten Weltkrieg erfahrungsgeschichtlich untersuchen will, braucht man Interviews mit Menschen, die zum großen Teil schon jetzt nicht mehr leben. Entsprechend wichtig werden Interviews, die in der Vergangenheit geführt wurden und in Archiven zugänglich sind. Ein Hindernis für solche Sekundäranalyse von Interviews ist zu meist, dass die Auswertung sehr zeitaufwändig ist. Da ist der online-Zugang ein echter Innovationsschub.

Was wir seit einigen Jahren erleben, ist die Verwandlung mündlicher historischer Quellen in audiovisuelle Forschungsdaten. Als Interviewarchiv kommen wir gar nicht daran vorbei, diesen Weg mitzugehen. Und ich tue das mit Überzeugung, weil ich es für sinnvoll halte, diesen Prozess mitzugestalten. Archivierung und Forschung können und werden von den digitalen Möglichkeiten profitieren. In meiner Verantwortung für das Archiv „Deutsches Gedächtnis“ suche ich aktiv die Zusammenarbeit mit Kolleg:innen aus der Informatik.⁵ Den Prozess mitzugestalten, heißt aber auch, die Implikationen der Digitalisierung kritisch im Auge zu behalten. Dabei sehe ich heuristische, epistemologische und ethische Implikationen.

Digitalisierung und biographische Forschung – Fluch und Segen

Heuristik meint in der Geschichtswissenschaft die Bestimmung und das Auffinden von Quellen, die geeignet sind, die jeweilige Forschungsfrage zu beantworten – oder vielleicht bescheidener: zu bearbeiten. Die Digitalisierung und das Internet haben diesen Prozess radikal verändert. Das gilt natürlich auch für biographische Interviews. Dass im Archiv „Deutsches Gedächtnis“ 3.000 Oral History-Interviews für Forschungen bereitstehen, findet man heute schnell heraus. Und das ist gut so. Problematisch wird es, wenn sich die weitere Suche auf das beschränkt, was digital zugänglich ist. Was online verfügbar ist, wird gern genutzt. Was nur im Archiv eingesehen werden kann, bleibt außerhalb der Wahrnehmung. Die Digitalisierung führt zu einer Verzerrung im Wettbewerb um Aufmerksamkeit. Man kann sich leicht ausmalen, wie dies sachfremde Ausschluss-

mechanismen und Aufmerksamkeitsfilter produziert, die Forschungsprozesse unter der Hand mitbestimmen. Und zwar in der Weise, dass sich die Auswahl der Quellen nicht danach ausrichtet, was zur Bearbeitung einer Frage am aussichtsreichsten beiträgt, sondern was bequem genutzt werden kann. Meine Sorge ist, dass das Nicht-Digitale dem kulturellen Vergessen anheimfällt, aber auch weitergehend, dass die Suche mehr und mehr den Regeln des Internets folgt, wo man vor allem das findet, was schon andere gefunden haben, so dass die Trefferliste zur Rangliste wird, bei der die Interviews mit den meisten „Klicks“ und „Likes“ ganz oben stehen.

Unser Archivportal *Oral History.digital* ist zweifellos ein gutes Angebot. Dass ein Archivbesuch online ohne Beratung auskommen muss, halte ich allerdings für eine bedenkliche Begleiterscheinung. Tatsächlich hat der Mensch in dem Mensch-Maschine-System Archiv, die Archivarin bzw. der Archivar, eine wichtige Funktion. Nach meiner Erfahrung entstehen oftmals erst im Gespräch zwischen Forschendem und Archivar:in gute Suchstrategien, um mit den vorhandenen Findmitteln – Metadaten und Volltextsuche – relevante Interviews zu ermitteln. Im Online-Archiv gibt es statt Beratung vorgefertigte Register, die sich an den erwartbaren Rechercheanfragen orientieren. Da bleibt wenig Raum für Kreativität und forschende Phantasie, was zu den epistemologischen Implikationen überleitet, die auch auf ethische Aspekte ausstrahlen.

Den timecodierten Transkripten verdanken wir die synchrone Darstellung von Audio bzw. Video und Transkript und damit die Erweiterung der Interpretation vom gedruckten Text hin zur Sprache in Ton und Bild. Gleichzeitig bietet diese Technologie die Möglichkeit, die Relevanz der Treffer unmittelbar zu überprüfen, indem man an genau die Stelle im Interview springt. Passt die Sequenz nicht zur Suche, wird der Treffer verworfen. Und weiter geht es mit dem nächsten Treffer in der Liste. Ein respektvoller Umgang mit denjenigen, die sich für ein lebensgeschichtliches Interview zur Verfügung gestellt haben und ihre Erinnerungen ungeschützt preisgegeben haben, sieht anders aus.

Das trifft natürlich auch, ja sogar noch deutlicher zu, wenn biographische Interviews zu Trainingsdaten für die Verbesserung der Spracherkennung werden. Hier wird völlig von jedem Inhalt, jeder Bedeutung abgesehen. Hier geht es allein um die Eigenschaft der Maschinenlesbarkeit, die der Soziologe Armin Nassehi als eine der wichtigsten Eigenschaften von Daten ausmacht.⁶ Die Datenförmigkeit erlaubt es, letztlich alles mit allem zu vergleichen. Der Begriff

4 Vgl. Oral-History.Digital. Informationsinfrastruktur für die Erschließung, Recherche und Annotation von audiovisuellen narrativen Interviews, <https://www.oral-history.digital/>.

5 Munir Salman/Felix Engel/Almut Leh/Matthias Hemmje, Informationstechnologische Unterstützung der Archivierung biographischer Interviews und Erinnerungszeugnisse, in: BIOS. Zeitschrift für Biographieforschung, Oral History und Lebensverlaufsanalysen, 30 (2017), S. 92–100, <https://doi.org/10.3224/bios.v30i1-2.08>.

6 Armin Nassehi, *Muster: Theorie der digitalen Gesellschaft*, 3. Aufl., München 2019.

„Kriegerwitwensohn“ steht dann nicht mehr für eine generationenspezifische biographische Erfahrung oder für das kommunikative Angebot einer Selbstdeutung. Der Begriff ist lediglich Anlass für die Frage, wie man dem Sprachmodell Komposita beibringen kann. Solcherart Gegenstand wissenschaftlicher Verwertungsinteressen zu werden, haben sich die Interviewten wohl kaum vorgestellt, als sie der Archivierung ihres Interviews zugestimmt haben.

Doch zurück zum Suchalgorithmus und zur Trefferauswertung. Nicht nur ethisch, sondern auch mit Blick auf die Forschung ist solch ein selektiver, auf Ausschnitte beschränkter Zugang problematisch, weil er das Erkenntnispotential biographischer Interviews beinahe vollständig verfehlt. Selten geht es in den Interviews um Faktisches, das in einem kurzen Satz Niederschlag finden könnte. Fast immer geht es um größere Zusammenhänge wie die Aufschichtung von Erfahrung im biographischen Verlauf oder um die Konstituierung von Sinn durch die Erzählung. Dimensionen, deren Analyse auf dem gesamten Interview fußt.

Oral History als Forschungsmethode ist ein zutiefst induktives Verfahren, bei dem der Forschungsprozess durch ein hohes Maß an Offenheit gekennzeichnet ist. Eben deshalb werden die Interviews so geführt, dass sie dem Interviewten einen Erzählraum öffnen, den dieser/diese nach eigenen Relevanzkriterien füllen und den Spuren seines Gedächtnisses ungehindert folgen kann. Das bringt mich zum nächsten Punkt.

Bei der Recherche mittels Volltextsuche oder Keywords geht das Unausgesprochene verloren. Wir alle wissen, dass man intensiv über Dinge sprechen kann, ohne sie auf den Begriff zu bringen. Hier könnte eine Wortfeldanalyse vielleicht weiterhelfen. Aber was ist mit all dem, über das gar nicht gesprochen wird? Was verschwiegen wird, kann manchmal wichtiger sein, als das, was ausführlich berichtet wird. Tatsächlich sind es oftmals gerade die Brüche und Lücken in einer Erzählung, die den Schlüssel zum Sinn-Verstehen liefern. Die kann man nur aufspüren, wenn man das gesamte Interview analysiert.

Nach meiner Beobachtung steht die Sekundäranalyse von Oral History-Interviews in der Gefahr, die biographischen Erzählungen als Steinbruch zu nutzen, aus dem die Zitate herausgebrochen werden, die zu den Ergebnissen passen, die aus anderen Quellen gewonnen wurden.⁷ Diese Tendenz hat es immer gegeben, die elektronischen Rechercheinstrumente verstärken aber diese Praxis. Das geht soweit, dass die Forschungsfragen an die Antwortmöglichkeiten der Suchmaschine angepasst werden, mit der Folge, dass die Forschung kaum neue Erkenntnisse hervorbringen kann. Die Interviews bestätigen am Ende das, was zur These passt. Gemessen am Potential der Quelle ist dies eine unglaubliche Verflachung und Beschränkung.

Computerprogramme können diese Erkenntnisreduktion nochmals dramatisch verstärken, dann nämlich, wenn künstliche Intelligenz ins Spiel kommt. Beim maschinellen Lernen kommt es immer wieder zu problematischen Phä-

nomenen, die als Bias der Algorithmen bezeichnet werden. Lernende Algorithmen verstärken die Muster, die sie in den Trainingsdaten aufspüren. Beispiele dafür sind inzwischen zahlreich. Wenn ein Algorithmus, der Bewerbungen vorsortieren soll, mit den erfolgreichen Bewerbungen vergangener Jahre trainiert wird, ist die Wahrscheinlichkeit groß, dass das Kriterium „Männlich“ als Eigenschaft eines erfolgreichen Bewerbers qualifiziert wird. Das System wird deshalb bevorzugt männliche Bewerber vorschlagen. Lernende Algorithmen neigen dazu, den Status quo zu verfestigen oder noch zu verstärken.

Die Wissenschaftsjournalistin Manuela Lenzen hat dies schön beschrieben: „Kein Algorithmus ist klug genug, um zu verstehen, dass bestimmte Entscheidungen für bestimmte Bevölkerungsgruppen diskriminierend sind. Und kein Algorithmus ist klug genug, um zu verstehen, dass wir uns eine Zukunft wünschen, die sich von der Vergangenheit unterscheidet.“⁸ Ich fürchte, dass es mit der Klugheit von Algorithmen auch dann nicht weit her ist, wenn es darum geht, Bedeutung und Sinn in biographischen Narrationen zu erkennen, wenn das Ziel nicht darin besteht, auf das immer Gleiche zu verweisen, sondern Veränderungen zu erkennen.

Perspektiven einer digitalen Hermeneutik

Die Transformation von Interviews in maschinenlesbare Daten betont strukturelle Ähnlichkeit und macht Unterschiedliches vergleichbar. Computerprogramme sind auf Muster trainiert. Was nicht in das Schema des Musters passt, wird ausgesondert. Für das hermeneutische Verstehen der biographischen Sinnkonstruktion kann aber gerade das wichtig sein, was durch das Muster hindurchfällt. Ich will gar nicht leugnen, dass auch computergestützte Analysen zu geschichtswissenschaftlich relevanten Erkenntnissen führen können. Ich bin sogar gespannt darauf. Die Herausforderung besteht aber meines Erachtens darin, computergestützte Verfahren so zu gestalten, dass sie die Charakteristika historischer Forschung, induktives Vorgehen und hermeneutisches Verstehen, nicht behindern oder gar verhindern, sondern unterstützen.

Beide Seiten, Historiker:innen wie Informatiker:innen, können von diesem Prozess auch für ihre eigene Profession profitieren. Historiker:innen, indem sie versuchen, ihr intuitives Vorgehen an Regeln zu binden; Informatiker:innen, indem sie Unschärfe zulassen. Für die Geschichtswissenschaft lautet die spannende Frage: Wie kann man hermeneutische Prozesse besser modellhaft darstellen – und damit überprüfen? Für die Informatik besteht die Herausforderung und zugleich die Chance darin, ihre lernenden Systeme flexibler zu machen. Vielleicht können Algorithmen dann doch Hermeneutik. Und vielleicht muss sich ein

7 Vgl. Linde Apel, Oral History reloaded. Zur Zweitauswertung von mündlichen Quellen, in: Westfälische Forschungen. Zeitschrift des LWL-Instituts für westfälische Regionalgeschichte, 65 (2015), S. 243–254.

8 Manuela Lenzen, Künstliche Intelligenz. Fakten, Chancen, Risiken, München 2020, S. 53.

solches Mensch-Maschine-System nicht in Fragen und Antworten erschöpfen, sondern öffnet Räume für Kreativität und Neugier.

Bis es so weit ist, sollten wir darauf achten, dass wir uns nicht von der Maschine die Fragen vorgeben lassen, indem wir nur noch Fragen stellen, die die Maschine beantworten kann. Wir brauchen nicht die Maschine, die uns die Frage nennt, auf die die Antwort 42 ist. Die Grenzen der

Forschung sollten durch das Denken bestimmt sein, nicht durch das maschinell Mögliche. ■



Dr. Almut Leh
Institut für Geschichte und Biographie,
FernUniversität Hagen
almut.leh@fernuni-hagen.de

(Foto: © Sibylle Ostermann)