Forschungsinstitut Betriebliche Bildung

f-bb-online

Dr . Andreas Fischer

Die Zukunft der Qualitativen Inhaltsanalyse im Zeitalter Künstlicher Intelligenz

Ein Ablaufmodell KI-basierter Inhaltsanalyse

Impressum

f-bb-online
Schriftenreihe des Forschungsinstituts Betriebliche Bildung (f-bb)
ISSN 2197-8026

Herausgegeben von

Dr. Iris Pfeiffer
Forschungsinstitut Betriebliche Bildung (f-bb) gGmbH
Rollnerstraße 14
90408 Nürnberg
www.f-bb.de

Das Forschungsinstitut Betriebliche Bildung (f-bb) arbeitet seit 2003 an der Weiterentwicklung des Systems der beruflichen Bildung durch Forschung in Deutschland und international. Das Leistungsspektrum umfasst die Durchführung von Modellversuchen, Gestaltungs- und Transferprojekten, die wissenschaftliche Begleitung von Förderprogrammen, die Evaluation von Verordnungen und und Maßnahmen sowie die Umsetzung von Fallstudien, empirischen Erhebungen und Analysen.

Autor

Dr. Andreas Fischer

Erscheinungsjahr

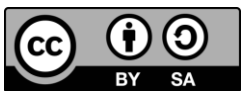
2025

Diese Publikation ist frei verfügbar zum Download
unter www.f-bb.de

Zitiervorschlag

Fischer, A. (2025). Die Zukunft der Qualitativen Inhaltsanalyse im Zeitalter Künstlicher Intelligenz. Ein Ablaufmodell KI-basierter Inhaltsanalyse. f-bb-online 03/25.

Diese Publikation ist unter folgender Creative-Commons-Lizenz veröffentlicht:



Inhalt

Inhalt 3

1. Die Nuancen verstehen: Eine methodologische Untersuchung der KI-gestützten qualitativen Inhaltsanalyse 4

2. Eine kurze Einführung in QIA 7

2.1. KI-basierte Inhaltsanalyse 9

3. Ein paar Worte zu KI-Kenntnissen 12

3.1. Prompt Prinzipien für KI-basierte Inhaltsanalyse 14

4. Ablaufmodell KI-basierter Inhaltsanalyse 16

1. Kontext 17

2. Technik 17

3. Einheiten 17

4. Implementierung 20

5. Evaluation 21

5. Beispiele KI-basierter Inhaltsanalysen 22

5.1. KI-basierte Zusammenfassung 22

5.2. KI-basierte Explikation 24

5.3. KI-basierte Induktion 25

5.4. KI-basierte Deduktion 27

6. Diskussion 30

7. Literatur 32

Außerdem zuletzt vom f-bb veröffentlicht 35

1. Die Nuancen verstehen: Eine methodologische Untersuchung der KI-gestützten qualitativen Inhaltsanalyse

Die Qualitative Inhaltsanalyse (QIA), die traditionell auf menschlicher Expertise basiert (Berelson, 1952), galt lange Zeit als etwas, das Maschinen nie in der Lage sein würden zu leisten (Weizenbaum, 1977; Döring & Bortz, 2016, 608). Fortschritte in der natürlichen Sprachverarbeitung (NLP, Bittermann & Fischer, 2024), insbesondere im Bereich generativer künstlicher Intelligenz (KI), wie ChatGPT und Transformer-Modelle im Allgemeinen (Vaswani et al., 2017), lösten diese Vorbehalte jedoch in den letzten Jahren zunehmend auf: Large Language Models (LLMs) und andere KI-Systeme entwickeln zunehmend Fähigkeiten in der qualitativen Forschung, die an menschliche Expertise heranreichen (Fischer et al., 2023; Fischer, im Druck) und es gibt eine wachsende Zahl von Vorträgen und Schulungen von professionell Forschenden über den Einsatz von KI in der qualitativen Forschung (z. B. Friese et al., 2024).

Entscheidend für die Verbreitung von KI in Forschungskontexten ist auch, dass Open-Source-KI-Systeme oder aktuelle KI-Modelle mit offenen Gewichten (engl. *open-weights models*), wie Googles Gemma, Metas Llama, Microsofts Phi oder Mistrals Mixtral (sowie deren Nachfolger, Varianten und verschiedene sog. *„Finetunings“*) eine digital souveräne Anwendung von KI-basierter Inhaltsanalyse ermöglichen - d.h. Forschenden eine (kosten)freie und datenschutzkonforme Analyse auf ihrer eigenen Infrastruktur ermöglichen (siehe Fischer et al., 2023; Fischer et al., 2024a).

KI-gestützte qualitative Inhaltsanalyse im Allgemeinen bezeichnet eine wissenschaftliche Methode, um gestützt auf moderne KI sowie interpretativ, nicht ausschließlich quantifizierend, Kommunikationsinhalte hinsichtlich ihrer Bestandteile zu untersuchen.

Inhaltsanalyst*innen haben heutzutage eine große Bandbreite an Möglichkeiten, KI in ihren Analyseprozess zu integrieren: Sie können KI beispielsweise nutzen, um Materialien in einem freien Dialog mit einem KI-Chatbot zu diskutieren, oder sie können einen systematischen Ansatz wählen und der KI ihrer Wahl ausgewählte Fragen auf standardisierte Weise stellen - entweder über eine grafische Benutzeroberfläche (GUI) oder auf automatisierte Weise über eine Anwendungsprogrammierschnittstelle (API) oder eine Befehlszeilenschnittstelle (CLI). Diese beiden Ansätze der Interaktion mit der KI entsprechen dem, was Dell'Acqua et al. (2025) ikonisch als *„Cyborg“*- und *„Centaur“*-Modus der Arbeit mit KI bezeichnet haben:

- der unsystematische Chat mit der KI entspricht dem *„Cyborg“*-Modus – im Sinne einer vollständigen Integration des eigenen Aufgabenablaufs mit der KI, in ständiger Interaktion mit ihr.
- der systematische Ansatz entspricht dem *„Centaur“*-Modus - im Sinne einer klaren Aufteilung und Delegation von Teilen und Schritten des Arbeitsablaufs an die KI oder an sich selbst (vgl. Dell'Acqua et al., 2025).

Generell sind die Möglichkeiten der Nutzung von KI-Modellen zur Analyse von Inhalten aller Art inzwischen weit verbreitet und in verschiedenen Formen kostenlos verfügbar:

- Zum einen gibt es eine Vielzahl von freien und quelloffenen KI-Systemen, die zur Analyse von Textkorpora und anderen Materialien sowohl online (z.B. Huggingface Chat UI) als auch offline und auf eigener Infrastruktur (z.B. Jan.ai) einfach eingesetzt werden können. Auch etablierte QIA-Tools wie ATLAS.ti (Friese, 2019) oder MaxQDA (Kuckartz & Rädiker, 2019) verfügen über auf moderner KI basierende Funktionen bzw. Addons, die die Datenanalyse mit Funktionen wie KI-basiertem Chat über Materialien oder Zusammenfassungen, Codegenerierung oder Codezuweisung unterstützen.
- Auf der anderen Seite werden Programmiersprachen wie Python oder R - sowie spezielle Pakete wie topiclabels von Peters et al. (2024) oder aifeducation von Berding et al. (2025) - von Datenwissenschaftler*innen und -analyst*innen zur Datenanalyse verwendet und ermöglichen komplexe und flexible Arbeitsabläufe, die KI-Systeme zur Zusammenfassung, Explikation, Induktion und Deduktion nahtlos in einer kontrollierten, potenziell replizierbaren und digital souveränen Art und Weise auf der Grundlage freier Open-Source-Software integrieren.

Das Potenzial automatischer Zusammenfassungen, Explikationen, Induktionen und Deduktionen auf der Grundlage generativer KI scheint unbestreitbar, aber ein wissenschaftlicher Ansatz für KI-gestützte QIA ist auch in der heutigen Zeit noch nicht trivial und erfordert eine sorgfältige Navigation durch den Raum des Möglichen. Der vorliegende methodologische Beitrag stellt daher die *KI-basierte Inhaltsanalyse* nach Fischer (im Druck) vor, und bietet damit ein strukturiertes Rahmenwerk in dieser sich entwickelnden Forschungslandschaft. Dieses Rahmenwerk soll dabei helfen, über vereinfachende Anwendungen hinauszugehen und zu einem robusten und validen KI-gestützten QIA-Ansatz anleiten.

Hinweise für die Praxis

Für den unkomplizierten Einstieg in die Nutzung von KI on premises gibt es nutzerfreundliche Produkte mit grafischer Benutzerschnittstelle (GUI), wie beispielsweise LMStudio (für den persönlichen nicht-kommerziellen Gebrauch) oder Jan.ai (unter freier Lizenz), die neben dem Chat mit Text-generierender KI auch die Möglichkeit bieten, die KI über einen http-Server anderen Programmen im eigenen Netzwerk zugänglich zu machen (sodass man sie z.B. über Programmiersprachen wie Python oder R in klassische Datenauswertungen integrieren kann). Für den professionellen Einsatz auf dezidierten KI-Servern empfehlen sich hingegen Softwarebibliotheken wie llama.cpp oder ollama, die dann primär über eine Programmierschnittstelle (API) in eigene Softwareprodukte und/oder Arbeitsabläufe eingebunden werden können.

Wir plädieren für einen systematischen Weg der Analyse qualitativer Inhalte mit KI, der eine iterative Verarbeitung des Materials durch eine Reihe von Anfragen beinhaltet (z. B. für eine induktive Kategorienbildung mit Hilfe von Textgenerierung oder für eine deduktive Kategorisierung auf der Grundlage robuster Texteinbettungen, vgl. Palomino et al., 2024). Das Rahmenwerk von Fischer (im Druck) betont jedoch grundsätzlich auch das *Mitdenken des Menschen*

in jeder Phase der Analyse - von der Planung und Formulierung der ersten Prompts bis hin zur Überprüfung der Prozesse und der abschließenden Validierung der Ergebnisse.

Nur durch ein systematisches, überwachtes und iteratives Zusammenspiel lässt sich das volle Potenzial der KI nutzen und gleichzeitig die der qualitativen Forschung innewohnende Stringenz und Vertrauenswürdigkeit gewährleisten. Dieser Beitrag befasst sich sowohl mit praktischen Implementierungen als auch mit methodologischen Überlegungen, und ebnet damit den Weg für eine zuverlässigere und aufschlussreichere KI-gestützte QIA der Zukunft.

2. Eine kurze Einführung in QIA

Die qualitative Inhaltsanalyse bezieht sich auf eine systematische und interpretative Untersuchung von Kommunikationsinhalten, wobei diese Inhalte typischerweise nicht-numerischer Natur sind, wie z. B. Texte. Bei dieser Art der Analyse wird der Inhalt systematisch in seine Bestandteile zerlegt, um seine Zusammensetzung, Themen und Bedeutungen zu verstehen, zu interpretieren oder Schlussfolgerungen zu ziehen.

Was die qualitative Inhaltsanalyse von anderen Formen der Inhaltsanalyse unterscheidet, ist, dass sie über die bloße Quantifizierung von Daten, wie z. B. das Zählen von Begriffshäufigkeiten, hinausgeht und sich stattdessen auf ein tieferes, nuanciertes Verständnis der Bedeutung, des Kontexts und der zugrunde liegenden Botschaften des Inhalts konzentriert. Wie Kuckartz & Rädiker (2023) es ausdrücken: *„Qualitative Inhaltsanalyse ist die systematische und methodisch kontrollierte wissenschaftliche Analyse von Texten, Bildern, Filmen und anderen Inhalten der Kommunikation“* (Kuckartz & Rädiker, 2023, 21, übersetzt). Mit Blick auf die Praxis des inhaltsanalytischen Arbeitens scheint erwähnenswert, dass die qualitative Inhaltsanalyse durchaus auch eine Quantifizierung nicht-numerischer Daten umfassen kann (aber nicht darauf beschränkt sein darf) - man denke beispielsweise an den üblichen Prozess des Codierens und anschließenden Auszählens der Nennungshäufigkeiten von Themen oder Konzepten, die sich den Daten entnehmen lassen (z. B. Mayring, 2022). Darüber hinaus kann die qualitative Inhaltsanalyse grundsätzlich verschiedenen Zwecken dienen, darunter beispielsweise die Sondierung, die Theoriebildung und die Theorieprüfung. Obwohl die Theorieprüfung häufig eher mit quantitativen Forschungstraditionen in Verbindung gebracht wird (z. B. Döring & Bortz, 2016), können qualitative Ansätze auch für gewisse Spielarten der Prüfung von Theorien eingesetzt werden, was die Flexibilität und Vielseitigkeit von qualitativen Analysen unterstreicht (vgl. Strauß & Corbin, 1998).

Hinweise für die Praxis

Dass sich qualitative Methoden und speziell auch Inhaltsanalysen neben der Theoriebildung auch für gewisse Spielarten der Theorieprüfung einsetzen lassen zeigen insb. Studien zum Thema *„Grounded Theory“*.
In diesem Zusammenhang kommt beispielsweise die *Zusammenfassung* von Passagen zu paraphrasierenden Codes für Konzepte, sowie die *Induktion* von theoretisch bedeutsamen Kategorien (sog. *„Offenes Codieren“*) ebenso zum Einsatz wie die *Explikation* von Zusammenhängen zwischen Kategorien (sog. *„Axiales Codieren“*) oder die *Deduktion* von Implikationen zur Selektion, Überprüfung und Anpassung von zentralen Elementen des Kategoriensystems (sog. *„Selektives Codieren“*) – bei Interesse an Details zur Methodologie sei beispielsweise verwiesen auf vgl. Strauß & Corbin (1998) sowie Glaser & Strauß (1967).

Wissenschaftliche Theorien beruhen auf theoretischen Konstrukten und Konzepten (Kategorien) sowie deren Zusammenhängen. Wie Döring und Bortz (2016, 222f.) darlegen, können Kategorien entweder aus dem Datenmaterial gewonnen werden, indem vom Spezifischen auf das Allgemeine geschlossen wird (sog. *Induktion*) oder aber anhand von theoretischen Definitionen und Ableitungen spezifiziert und/oder auf das Datenmaterial angewendet werden, indem vom Allgemeinen auf das Spezifische geschlossen wird (sog. *Deduktion*). Auch deskriptive Zusammenfassungen ohne dedizierten Theoriebezug (sog. *Zusammenfassung*) sind in der Literatur weit verbreitet (z. B. Fischer et al., 2022), ebenso wie Explikationen von Passagen auf der Basis von Zusatzmaterialien (sog. *Explikation*) - oft in Kombination mit anderen Auswertungsmethoden und als Vorstufe zu einer theoriegenerierenden Analyse (z. B. Ahrbeck et al., 2009, 105f).

Die qualitative Inhaltsanalyse zielt typischerweise darauf ab, aus qualitativem Text- oder Bildmaterial (o.ä.) systematisch Inhalte zu extrahieren, indem letztlich Kategorien gebildet werden und diese ggf. quantifiziert werden (Döring und Bortz, 2016, 602). Die Inhaltsanalyse im Allgemeinen hat ihren Ursprung in den Kommunikationswissenschaften (z.B. Berelson, 1952; vgl. Mayring, 2016), als *„die Anwendung reproduzierbarer und valider Methoden, um von einem Text auf andere Zustände oder Eigenschaften seiner Quelle zu schließen“* (Krippendorf, 1969, 103, übersetzt), meist umgesetzt als einfaches Auszählen von Wörtern (*„quantitative Inhaltsanalyse“*, Mayring, 2016). Später wurde sie auf verschiedene Weise weiterentwickelt, um latente Inhalte einzubeziehen, was neben modernen Formen der quantitativen Inhaltsanalyse (z. B. Fischer, Hilse, Schütt-Sayed, 2023; Bergmann et al., 2024; Fischer & Pfeiffer, 2019) zur *„qualitativen Inhaltsanalyse“* führte, wie oben beschrieben.

In der Regel basiert die qualitative Inhaltsanalyse auf einer Strategie nach dem Motto *„teile und herrsche“*, die für viele Arten der Big-Data-Analyse geeignet ist - d. h. zur Analyse eines Materialkorpus arbeiten sich Analyst*innen iterativ durch das Material (sog. *„Verarbeitungseinheiten“* oder, wie Mayring, 2022, sie nennt, *„Auswertungseinheiten“*) und aktualisieren die Analyse nach jedem Verarbeitungsschritt. Traditionell wird diese Art der Analyse meist von Menschen durchgeführt (obwohl sie zunehmend durch technologische Hilfsmittel unterstützt wird, z. B. durch spezielle Computerprogramme, siehe Mayring, 2022). Wenn die Einheiten unabhängig voneinander verarbeitet werden, kann die Analyse jeder Einheit parallel erfolgen (durch andere Analyst*innen oder andere Server) und anschließend können die Ergebnisse aller Teil-Analysen für eine skalierbare Verarbeitung integriert werden, was in der Big-Data-Analyse als *Map-Reduce-Ansatz* bekannt ist (Marz & Warren, 2016).

Mayring (1985) entwickelte eine Reihe von Verfahren und Techniken zur systematischen Inhaltsanalyse, um die Vorteile der quantitativen Inhaltsanalyse zu erhalten und sie auf qualitativ-interpretative Schritte zu übertragen und weiterzuentwickeln (Mayring, 2000), indem er u.a. den kommunikativen Kontext von Materialien, schrittweises Vorgehen, Kategorien als Zentrum der Analyse sowie Kriterien wie Reliabilität und Validität betonte (ebd.).

Ausgehend von drei Grundformen der Interpretation hat Mayring (1985) die *zusammenfassende* Inhaltsanalyse, die *explizierende* Inhaltsanalyse und die *strukturierende* Inhaltsanalyse entwickelt. Später hat er auch die Unterscheidung zwischen induktiver Kategorienbildung und deduktiver Kategorienzuordnung genauer herausgearbeitet (Mayring, 2000; Mayring, 2016; Mayring, 2022) und versucht, sie in einen gemeinsamen Rahmen zu integrieren (Mayring, 2016; Mayring, 2022). So unterscheidet Mayring (2016), ausgehend von der ersten Grundform der Interpretation (*„Zusammenfassen"*), (a) die *„zusammenfassende Inhaltsanalyse"*, die versucht, das gesamte Material auf das Wesentliche zu reduzieren, und (b) die *„induktive Kategorienbildung"*, bei welcher nur bestimmte theoriebezogene Komponenten berücksichtigt bzw. aus dem Material abgeleitet werden. Weiterhin unterscheidet Mayring (2016) verschiedene Arten der *„strukturierenden Inhaltsanalyse"* als Varianten/Teilmengen der deduktiven Kategorienanwendung - nämlich *„inhaltliche"*, *„formale"*, *„typisierende"*, *„skalierende"/„evaluative"* Inhaltsanalyse. Insbesondere impliziert nach Mayring (2022, S.67) *„Strukturierung"* immer auch *„deduktive Kategorienanwendung"*, sodass sich mit Fischer (im Druck) ein Vierfelder-Schema der qualitativen Inhaltsanalyse ableiten lässt, wie im folgenden Unterkapitel näher ausgeführt wird. In ähnlicher Weise schlagen Kuckartz und Rädiker (2023) jeweils ein Prozessmodell für die Entwicklung von Kategorien und Unterkategorien zur Strukturierung der Daten (*„strukturierende QIA"*), für die Entwicklung evaluativer Kategorien und die Bestimmung unterschiedlicher Ausprägungen (*„evaluative QIA"*) oder für die Bildung und Zuordnung von Typen zu Fällen auf der Grundlage des Attributraums der Daten (*„typenbildende QIA"*) vor.

2.1. KI-basierte Inhaltsanalyse

Im vorliegenden Beitrag werden wir Zusammenfassung, Explikation, Induktion und Deduktion als die wichtigsten Arten und Grundtechniken der KI-basierten Inhaltsanalyse näher erläutern und dabei auch auf zusätzliche Varianten der QIA eingehen.

Wie Tabelle 1 zeigt, beziehen sich die induktive Kategorienbildung (*„KI-basierte Induktion"*) und die deduktive Kategorienanwendung (*„KI-basierte Deduktion"*) bei der KI-basierten Inhaltsanalyse explizit auf Kategorien und damit auf die Entwicklung bzw. Überprüfung wissenschaftlicher Theorien. Bei der zusammenfassenden Inhaltsanalyse (*„KI-basierte Zusammenfassung"*) und der explizierenden Inhaltsanalyse (*„KI-basierte Explikation"*) geht es hingegen um die Verdichtung respektive Erweiterung eines Materialkorpus – wobei bei Zusammenfassung und Explikation ein Bezug zu Kategorien nicht notwendig aber auch nicht ausgeschlossen ist (man denke z.B. an die zusammenfassende oder explizierende Analyse von Kategoriensystemen).

Table 1. Systematisierung KI-basierter Inhaltsanalyse

KI-basierte Zusammenfassung: KI wird eingesetzt, um eine Zusammenfassung auf der Grundlage eines Korpus spezifischer Materialien zu erstellen, unabhängig von theoretischen Kategorien.	**KI-basierte Explikation:** KI wird eingesetzt, um eine Explikation von Materialien innerhalb eines Korpus auf der Grundlage zusätzlicher Materialien/Kontexte/Assoziationen zu erstellen, unabhängig von theoretischen Kategorien.
KI-basierte Induktion: KI wird eingesetzt, um auf der Grundlage eines Korpus spezifischer Materialien Schlussfolgerungen zu theoretischen Kategorien und Konzepten zu ziehen.	**KI-basierte Deduktion:** KI wird eingesetzt, um auf der Grundlage einer Reihe theoretischer Kategorien und Konzepte Schlussfolgerungen zu bestimmten Materialien in einem Korpus zu ziehen.

Das in dieser Arbeit vorgeschlagene Ablaufmodell der KI-basierten Inhaltsanalyse nach Fischer (im Druck) ermöglicht die systematische Auswahl oder Kombination verschiedener Techniken der Inhaltsanalyse: So kann man z.B., wie in Tabelle 1 dargestellt, je nach Fragestellung die Texte eines Korpus im Hinblick auf eine Fragestellung zusammenfassen, Implikationen eines jeden Textes für die Fragestellung explizieren, Kategorien aus den Texten induzieren und/oder ableiten, zu welcher der induzierten Kategorien jeder Text gehört.

Darüber hinaus ermöglicht das im Folgenden vorgeschlagene Ablaufmodell nach Fischer (im Druck) die Integration verschiedener Arten von KI: So können beispielsweise KI-basierte Werkzeuge zur automatischen Spracherkennung für die Vor- oder Nachbearbeitung von Materialien (z.B. für die Transkription von Audiodateien) eingesetzt werden und auch im eigentlichen Analyseprozess gibt es eine Vielzahl von Methoden aus den Bereichen KI und NLP, die zusätzlich oder anstelle von textgenerierenden LLMs eingesetzt werden können (z.B. Bittermann & Fischer, 2024). Beispielsweise können traditionelle Ansätze zur Themenmodellierung (Topic Modeling), insbesondere in Kombination mit automatisierter Themenbenennung auf der Grundlage von LLMs (Peters et al., 2024), zur Realisierung einer Art KI-basierter induktiver Kategorienbildung eingesetzt werden (Bittermann & Fischer, 2018), und auch die Clusteranalyse traditioneller Texteinbettungen (Devlin et al., 2019) kann ein vielversprechender Ansatz sein, um eine KI-basierte Induktion von Kategorien auf der Grundlage von LLMs (aber ohne die sonst so charakteristische Texterzeugung) zu realisieren. Auch Retrieval-Augmented Generation (RAG) oder Embedding-based Retrieval (EBR) können für die Zuordnung von Kategorien zu Texten eingesetzt werden (vgl. Xu et al., 2021; Fischer, 2023; Fischer & Dörpinghaus, 2024; Berding et al., 2025; Palomino et al., 2024). All diese und viele weitere Ansätze können angewandt werden, um die Fähigkeiten von textgenerierenden LLMs auf die eine oder andere Weise zu ergänzen, und können innerhalb des vorgeschlagenen Ablaufmodells verwendet werden.

Wie Feuerriegel et al. (2025) es ausdrücken: *"Die explorative Inhaltsanalyse kann eine Reihe von NLP-Methoden umfassen. Zu den gängigen Ansätzen gehören frequenzbasierte Analysen (z. B. Zählen der am häufigsten verwendeten Begriffe), Koinzidenzanalysen (z. B. Untersuchung der Häufigkeit des gemeinsamen Auftretens von Wörtern), Named-Entity-Erkennung (z. B. Identifizierung von Personen, Organisationen oder Orten) und Clustering-Ansätze (z. B. Kategorisierung von Dokumenten in sinnvolle Themen). Visualisierung (z. B. Wortwolken) wird oft verwendet, um die extrahierten Muster besser erkennbar zu machen. Forscher verwenden oft eine Kombination von Methoden, um die Textdaten auf systematische Weise zu untersuchen"* (Feuerriegel et al., 2025, 97, übersetzt).

3. Ein paar Worte zu KI-Kenntnissen

Im Hinblick auf den allgemeinen Prozess der KI-basierten Inhaltsanalyse ist es wichtig zu beachten, dass KI im Sinne dieses Artikels trotz all ihrer Fähigkeiten nur ein Werkzeug ist, das von Analyst*innen verantwortungsvoll und kompetent eingesetzt werden sollte. Zu beachten ist, dass *„ein ausreichendes Maß an KI-Kenntnissen"* (KI-VO, Art. 4, übersetzt) eine grundlegende Voraussetzung für die verantwortungsvolle Anwendung von KI ist (vgl. Fischer et al., 2023) - dies gilt für KI-Systeme im Allgemeinen und für die KI-basierte Inhaltsanalyse im Besonderen.

Laut EU bedeuten „KI-Kenntnisse" (im Original *„AI-literacy"*, auch übersetzt mit dem konzeptuell umfassenderen Begriff *„KI-Kompetenz"*) ganz allgemein *„Fähigkeiten, Wissen und Verständnis, die es Anbietern, Anwendern und Betroffenen unter Berücksichtigung ihrer jeweiligen Rechte und Pflichten im Rahmen dieser Verordnung ermöglichen, KI-Systeme in Kenntnis der Sachlage einzusetzen sowie sich der Chancen und Risiken von KI und möglicher Schäden, die sie verursachen kann, bewusst zu werden"* (EU-KI-Gesetz, Art.3, Def.56, übersetzt) - doch was sind die Risiken von KI und die möglichen Schäden, die sie im Hinblick auf die Inhaltsanalyse verursachen kann? Ein vollständiger Überblick würde den Rahmen dieses Artikels sprengen, aber analog zu den Chancen möchten wir an dieser Stelle auch einen profunden Überblick über zentrale Risiken geben, die als Orientierung und Ausgangspunkt für eine bedarfsgerechte Vertiefung dienen können. Insbesondere bei der Arbeit mit LLMs zur Inhaltsanalyse ist zu beachten, dass es nicht trivial ist, optimale Ergebnisse zu erzielen, wenn man die Funktionsweise eines solchen Modells nicht versteht.

Hinweise für die Praxis

Analyst*innen, die KI für qualitative Inhaltsanalyse einsetzen, sollten als Betreiber von KI-Systemen (i.S.v. Artikel 3 KI-KVO) insb. auch ihre Rechte und Pflichten kennen. Dazu zählen Kenntnisse zu verbotenen Praktiken (Artikel 5 KI-VO) ebenso wie Transparenzpflichten (Artikel 50 KI-VO) uvm.

Hinsichtlich der Risiken und Schäden sollten Sie sich bewusst sein, dass die von LLMs erzeugten KI-Antworten...

- **... falsch sein können:** Jede von einem LLM erzeugte Antwort kann einen plausiblen Text enthalten, der falsch sein kann (eine sog. *„Halluzination"*), insbesondere wenn dem Prompt eine ausreichende Menge an Kontext fehlt, um die Frage zu beantworten. Um das Risiko zu verringern, falsche Schlüsse aus der Verwendung von KI-Antworten zu ziehen und Prompts angemessen zu testen, wählen Sie aktuelle, leistungsstarke Modelle aus und überprüfen Sie die generierten Antworten so gut Sie können.
- **... unfair sein können:** Die Ergebnisse eines LLM hängen stark von den Daten ab, mit denen es trainiert wurde, und daher können Modelle Stereotype reproduzieren, die in den Trainingsdaten häufig vorkommen. Diese Verzerrungen sind nicht immer explizit und offenkundig: Hofmann et al. (2024) haben zum Beispiel verdeckt rassistische Entscheidungen von LLMs auf Grundlage der Dialekte von Menschen hervorgerufen. Die Autoren zeigen zum Beispiel, dass verschiedene Sprachmodelle mit größerer Wahrscheinlichkeit annehmen, dass Sprecher eines „afroamerikanischen" Englisch eher weniger angesehene Berufe haben, sowie eher wegen Verbrechen verurteilt werden und zum Tode verurteilt werden (ebd.). Um das Risiko von Verzerrungen (engl. *biases*) zu verringern, kann es sinnvoll sein, die Daten so weit wie möglich zu anonymisieren, auf Verzerrungen zu achten und die Prompts und/oder die Modellauswahl gegebenenfalls anzupassen.
- **... nicht-deterministisch sein können:** Jede Anfrage an ein LLM kann zu unterschiedlichen Ergebnissen führen, insbesondere wenn der sog. *„Temperatur"*-Parameter nicht gleich Null ist. Um die Replizierbarkeit der Ergebnisse zu verbessern, sollte die Temperatur auf Null gesetzt werden. Wie Schlögl, Kupek und Böhme (2021) gezeigt haben, können die Vorhersagen neuronaler Netze zwischen verschiedenen CPU-Mikroarchitekturen variieren, und die Ergebnisse von Schlögl, Hofer und Böhme (2023) zeigen, dass die Dinge noch komplexer werden, wenn GPUs beteiligt sind. Wie Mühlhoff & Henningsen (2024) aufzeigten, können sich Bewertungen, die auf generativer KI basieren, aufgrund stochastischer Prozesse unterscheiden (obwohl sie für die große Mehrheit der getesteten Fälle nur geringfügige Abweichungen melden).
- **... von jedem Teil des Prompts abhängen können:** Jeder Teil des Prompts kann subtile Auswirkungen auf die Generierung des nachfolgenden Textes haben. Um mit dieser *„Prompt-Sensitivität"* der Ergebnisse umzugehen und verzerrte Ergebnisse zu vermeiden, kann es bei einigen Analysen ratsam sein, die Vorverarbeitung des Materials entsprechend zu ergänzen (z. B. Informationen zu Geschlecht, ethnischer Zugehörigkeit usw. umfassend zu entfernen, wenn Sie einzelne Arbeiten bewerten und sicherstellen wollen, dass diese Merkmale Ihre Analyse nicht verzerren). Stellen Sie insbesondere bei zustandsabhängigen Anfragen sicher, dass jeder Teil des Zustands (d. h. Informationen über die Historie früherer Verarbeitungsschritte) editierbar ist, und überwachen Sie jeden Schritt des Prozesses. Wenn Sie ein LLM über eine GUI oder eine Web-Schnittstelle (z. B. HuggingFace Spaces) verwenden und

eine Reihe von zustandslosen Anfragen (engl. *stateless requests*) emulieren wollen, sollten Sie sicherstellen, dass Sie für jede Anfrage einen neuen Chat starten.

- **... in Zukunft möglicherweise nicht mehr verfügbar sein können:** Für Analysen, die im Laufe der Zeit repliziert werden können, sollten Sie unbedingt ein sog. *„open-weights"* Modell verwenden und genau angeben, welche Version Sie verwendet haben. Auch Informationen über Hard- und Software können für eine möglichst genaue Replikation hilfreich sein. Bitte beachten Sie, dass sich die derzeit verfügbaren proprietären KI-Systeme im Laufe der Zeit ändern können (und sogar der Zugang eingeschränkt oder teurer werden kann), so dass die mit dieser Art von Systemen erzielten Ergebnisse möglicherweise nicht replizierbar sind.
- **... ressourcenintensiv generiert werden können:** Es wird empfohlen, Sprachmodelle zu verwenden, die (a) so klein wie möglich und (b) für die jeweilige Aufgabe geeignet sind (siehe unten für einige Empfehlungen, wo und wie Sie geeignete KI-Modelle für Ihre Analyse finden können), sowie die Token-Generierung - und die damit verbundene Zeit und Energie - auf ein Minimum zu beschränken. Sie können beispielsweise eine Menge Ressourcen sparen, wenn einfache Ja-Nein-Fragen zur binären Kategorienanwendung von kleinen Modellen durchgeführt werden, die sich darauf beschränken, ein einziges Token pro Frage zu generieren (z. B. *„Ja"* vs. *„Nein"*).

3.1. Prompt Prinzipien für KI-basierte Inhaltsanalyse

Obwohl es möglich ist, wertvolle Erkenntnisse aus der Antwort eines LLM zu gewinnen, indem man einfache Prompt-Templates wie *"Bitte fasse den folgenden Text zusammen: {Text}"* verwendet, können noch mehr (und vertrauenswürdigere bzw. validere) Erkenntnisse gewonnen werden, wenn ein gut gestalteter Arbeitsablauf angestrebt wird und Prompting-Prinzipien berücksichtigt werden. Enthält der Text selbst beispielsweise an Leser*innen gerichtete Aufforderungen (man denke an Stellenanzeigen, in denen Formulierungen wie *„Bewerben Sie sich jetzt!"*, *„Bitte geben Sie die folgenden Ziffern ein, um die Kontaktinformationen zu sehen"* usw. üblich sind), empfiehlt es sich im Allgemeinen, die eigenen Arbeits-Anweisungen eher an das Ende des Prompts zu stellen, um die Wahrscheinlichkeit zu erhöhen, dass der Dialog entsprechend fortgesetzt wird (z. B. *„Gegeben sei der folgende Text: \n \n{text}\n\nBitte erstelle eine Zusammenfassung!"*).

Als grundlegende Leitlinie für Prompt-Formulierungen, die sich an traditionelle LLMs richten, empfehlen wir die ROMANS-Strategie - ein Akronym aus den Begriffen *Role, Objective, Meta-Instructions, Application Examples, Notable Facts, Specified Recipients/Format* (in Anlehnung an die von Eggers, 2023, vorgeschlagene „ROMANE-Strategie", die sich auf die deutschen Begriffe *„Rolle"*, *‚Oberziel'*, *„Meta-Anweisungen"*, *„Anwendungsbeispiele"*, *„Nützliche Fakten"*, *„Empfänger"* bezieht) - mit einem optionalen A-Term. Anwendungsbeispiele sind z.B. von besonderer Bedeutung, um eine bestimmte Form der Ausgabe zu spezifizieren (z.B. eine bestimmte Art der Strukturierung der Ausgabe im JSON-Format), gleichzeitig können sie aber auch einen Anker setzen, der die Ergebnisse beeinflussen kann (vgl. *„anchoring heuristic"*, Furnham & Boo, 2011). Prompts mit Beispielen werden üblicherweise als *„single-shot"* oder *„few-shot"*

Prompting/Learning bezeichnet, während Prompts ohne Beispiele in der Regel der Standard sind und üblicherweise als *„zero-shot“* Prompting/Learning bezeichnet werden. Im Allgemeinen ist jede Komponente der ROMANS-Strategie im Prinzip optional und als Anregung für bedarfsgerechte Ergänzungen an Prompts gedacht, die nicht mit ausreichender Zuverlässigkeit die gewünschten Ergebnisse erzielen. Es können selbstverständlich auch andere Strategien und Prinzipien angewendet werden (für einen Überblick über empirisch validierte Prompting-Prinzipien siehe Bsharat, Myrzakhan, Shen, 2024).

Generell ist darauf zu achten, dass die Prompts entsprechend den Anforderungen des in der Analyse verwendeten KI-Systems und -Modells formatiert werden. Beispielsweise gibt es vortrainierte Modelle ohne Finetuning (bei denen das Prompt-Format weitgehend Ihnen überlassen bleibt) und Modelle, die speziell trainiert wurden für das Befolgen von Ein-Schritt-Anweisungen (sog. *„Instruction-Tuning“*), für das Fortführen von Mehr-Schritt-Dialogen (sog. *„Dialogue-Tuning“*) oder für das schrittweise Denken (sog. *„Reasoning-Tuning“*), und bei denen man gut beraten ist, Prompts im vorgesehenen Modell-spezifischen Format bereitzustellen (spezifiziert im sog. *„Prompt-Template“*). Einige KI-Systeme können Prompts auch automatisch in das gewünschte Format übersetzen, aber in jedem Fall müssen Sie sicherstellen, dass Sie die Prompts so bereitstellen, wie sie vom jeweiligen KI-System benötigt werden.

Im Allgemeinen hängen die Ergebnisse stark von dem für die Analyse verwendeten KI-Modell ab. Stellen Sie daher sicher, dass Sie ein für die jeweilige Aufgabe geeignetes Modell auswählen. Für eine Rangliste mit Informationen über die aktuellen Modelle in Bezug auf die Leistung in verschiedenen Sprachen (z.B. Deutsch) können Sie sich an Bestenlisten wie https://lmarena.ai/leaderboard wenden und das beste open-weights Modell auswählen, das mit der für die Analyse verfügbaren Hardware kompatibel ist. Die Leistung von KI-Modellen bei einem bestimmten Prompt hängt auch von den Modellparametern ab. Im Allgemeinen wird empfohlen, für jedes Modell Standardparameter zu verwenden (die in der Regel von den Modellautor*innen veröffentlicht werden), obwohl es einige Parameter gibt, die je nach Bedarf für die Analyse bewusst gewählt werden sollten: So gibt es beispielsweise den oben bereits erwähnten Parameter *„Temperatur“* (der die Höhe der zufälligen Abweichungen von der Generierung des wahrscheinlichsten Tokens angibt), der sich auf Null setzen lässt, um die wahrscheinlichsten Ergebnisse zu erhalten (siehe jedoch Schlögl, Hofer und Böhne, 2023). Zusätzlich ist es möglich, die Anzahl der generierten Tokens nach Belieben festzulegen (z. B. kann es sinnvoll sein, nur wenige Tokens zu generieren, wenn das KI-Modell für eine einfache Klassifizierungsaufgabe nur „ja“ oder „nein“ generieren soll. Manchmal kann auch ein einziges Token ausreichen, aber beachten Sie bitte, dass einige Modelle Sonderzeichen wie „*“ an den Beginn ihrer Antwort setzen können, sodass Sie ggf. etwas Puffer einplanen sollten).

4. Ablaufmodell KI-basierter Inhaltsanalyse

Bei der KI-basierten Inhaltsanalyse wird ein Korpus von Materialien (Texte, Bilder, Audios, Videos, etc.) mit Hilfe von KI-Systemen und -Modellen (insbesondere LLMs) analysiert, um eine bestimmte Forschungsfrage zu beantworten.

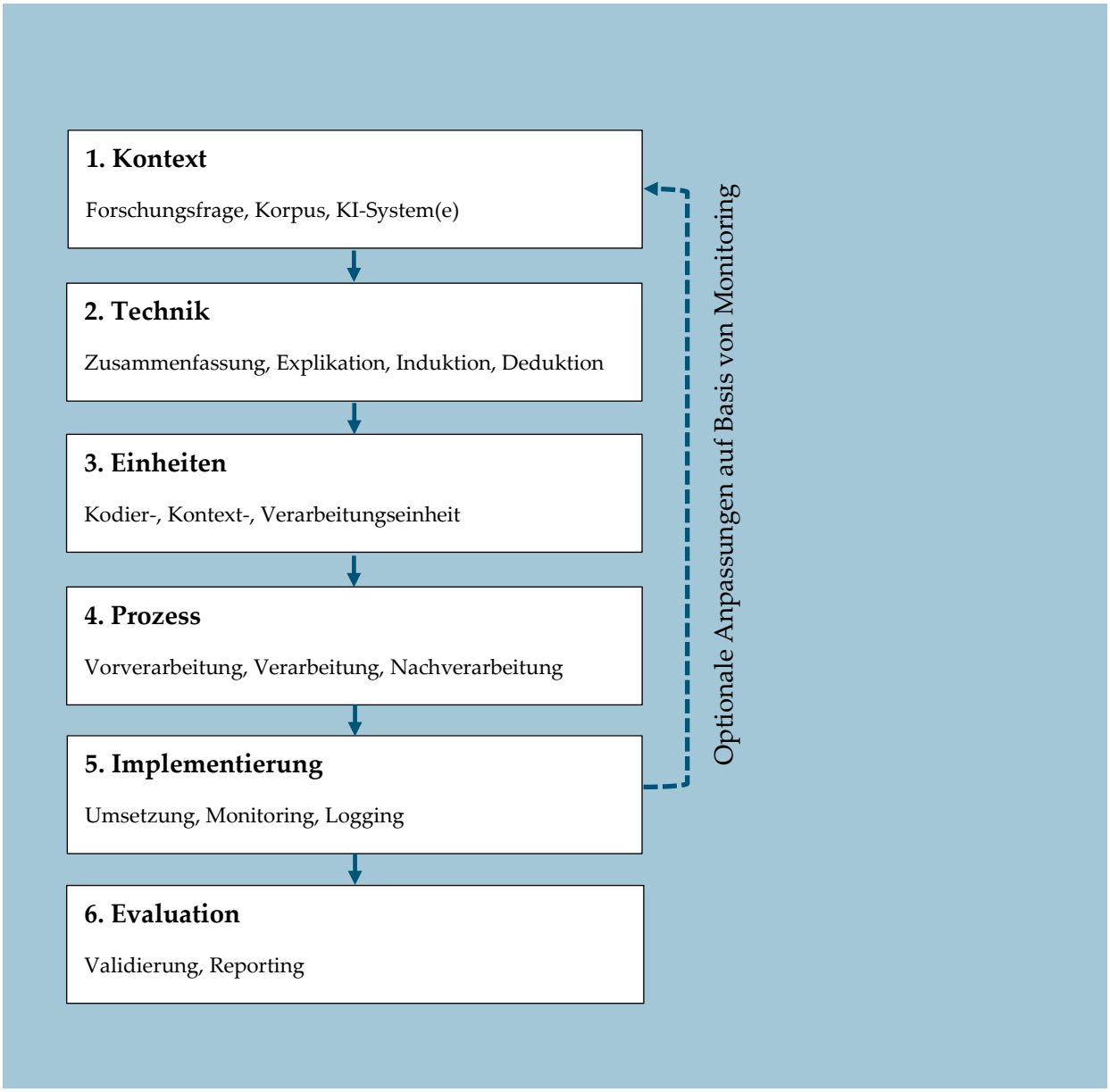


Abbildung 1. Ablaufmodell KI-basierter Inhaltsanalyse.

1. Kontext

Der erste Schritt der Analyse besteht in der Festlegung (a) einer Forschungsfrage, (b) eines Materialkorpus und (c) geeigneter KI-Systeme, Modelle oder Methoden zur Analyse des Korpus im Hinblick auf die Forschungsfrage (vgl. Mayring, 2022). Bei den KI-Systemen ist es grundsätzlich möglich, dedizierte textgenerative Modelle zu verwenden (und ggf. nicht-textuelle Materialien vorher zu transkribieren) oder multimodale KI-Systeme einzusetzen, die Multimedia-Input und/oder -Output verarbeiten können. Darüber hinaus können auch andere Ansätze aus den Bereichen KI oder NLP die Analyse unterstützen (vgl. Bittermann & Fischer, 2024). Im Folgenden konzentrieren wir uns auf textgenerative Modelle (LLMs) als Beispiele.

Grundsätzlich kann die KI-basierte Inhaltsanalyse durch Interaktion mit einem KI-System erfolgen im direkten Dialog (a) über eine grafische Benutzeroberfläche (wie z.B. ChatGPT, HuggingFace Spaces oder - lokal bzw. *on-premises* - LMStudio, Jan.ai usw.), oder automatisch (b) über eine API, oder (c) direkt aus einer CLI, die üblicherweise Python oder eine andere Programmiersprache verwendet. Die KI-Systeme selbst laufen in der Regel auf Servern (in der Regel mit leistungsstarken grafischen Verarbeitungseinheiten/GPUs und schnellen Prozessoren/CPUs), entweder on-premises oder in der Cloud.

2. Technik

Der zweite Schritt der Analyse besteht darin, eine geeignete Analysetechnik zu wählen, z. B. die KI-basierte Generierung von Zusammenfassungen (KI-basierte Zusammenfassung), Explikationen von Passagen (KI-basierte Explikation), induktive Kategorienbildung (KI-basierte Induktion), deduktive Kategorienanwendung (KI-basierte Deduktion) oder eine Kombination von Techniken (für Beispiele siehe unten). Im Falle der deduktiven Kategorienanwendung muss ein Kategoriensystem spezifiziert werden und im Allgemeinen sollte der theoretische Hintergrund der Studie recherchiert und angegeben werden, sofern nicht aus methodologischen Gründen ein theoriefreier Zugang angestrebt wird (z.B. Glaser & Strauß, 1967).

3. Einheiten

Der dritte Schritt ist die Festlegung der Analyseeinheiten (vgl. Mayring, 2022, Döring & Bortz, 2016, 603): Dabei ist in der Regel festzulegen, (a) welche Teile des Korpus separat kodiert bzw. besprochen werden können (Kodiereinheit), (b) welchem Material, Fall oder Set die jeweilige Einheit zuzurechnen ist (Kontexteinheit), und/oder (c) was als Verarbeitungseinheit(en) zu behandeln ist.

Hinsichtlich der Verarbeitungseinheit ist festzulegen, ob das gesamte Material in einem Verarbeitungsschritt (Massenverarbeitung, engl. *bulk processing*) oder durch Iteration über mehrere Verarbeitungseinheiten (schrittweise Verarbeitung, engl. *iterative processing*) verarbeitet werden soll, und - im letzteren Fall - (d) ob die einzelnen Verarbeitungseinheiten unabhängig voneinander verarbeitet werden sollen (zustandslose Anfragen, engl. *stateless requests*) oder ob sie aufeinander aufbauen (zustandshaltige Anfragen, engl. *stateful requests*). Bei der Festlegung der Analyseeinheiten für die KI-basierte Inhaltsanalyse sollten Einschränkungen der KI-Systeme (z.B. die Anzahl der Tokens, die in einer Anfrage

verarbeitet werden können - das sogenannte *„Kontextfenster“* - oder die Zeit, die für die Generierung einer Antwort benötigt wird) berücksichtigt werden.

- Die Massen-Verarbeitung ist am besten geeignet, wenn die Schlussfolgerungen zu jeder Kodiereinheit von den Schlussfolgerungen zu allen anderen Kodiereinheiten im Korpus abhängen können (z. B. wenn Sie eine globale Zusammenfassung aller Materialien im Korpus erstellen möchten) und wenn der Korpus klein genug ist, um auf diese Weise verarbeitet zu werden. Bitte beachten Sie, dass Kombinationen aus Massenverarbeitung und anderen Arten von Anfragen erforderlich sein können, um mit den Beschränkungen von KI-Systemen und/oder -Ressourcen fertig zu werden (für Einzelheiten siehe den Abschnitt zu Beispielen KI-basierter Inhaltsanalyse unten).
- Zustandslose Anfragen sind am besten geeignet, wenn die Schlussfolgerungen zu jeder Kodiereinheit nicht von den Schlussfolgerungen zu anderen Kodiereinheiten im Korpus abhängen sollen (z. B. wenn Sie jede Einheit einzeln auswerten wollen) und können auf Korpora beliebiger Größe angewendet werden.
- Zustandshaltige Anfragen sind am besten geeignet, wenn Schlussfolgerungen über jede Kodiereinheit von Schlussfolgerungen über alle zuvor analysierten Kodiereinheiten abhängen sollen (z.B. wenn Sie eine Liste von Kategorien kontinuierlich aktualisieren wollen, wenn neue Aspekte in einer Reihe von Texten erwähnt werden). Bei einer großen Anzahl von Token können zustandshaltige Anfragen als *„Retrieval Augmented Generation“* operationalisiert werden (siehe Xu et al., 2021; Fischer et al., 2024b), um Ressourcen zu sparen.

4. Prozess

Der vierte Schritt besteht darin, den Prozess, den jede Verarbeitungseinheit durchlaufen muss, zu spezifizieren und zu testen. Dies umfasst in der Regel (a) die Formulierung eines Prompts auf der Grundlage der Verarbeitungseinheit (Vorverarbeitung), (b) die Anforde-rung einer Antwort auf den Prompt von einem KI-System (Verarbeitung), (c) die Extraktion nützlicher Informationen aus der Antwort (Nachverarbeitung) für jede Verarbeitungseinheit.

- Zur Vorverarbeitung gehört in der Regel das Einfügen der Verarbeitungseinheit in ein Prompt-Template (z. B. *"Gegeben sei der folgende Text: {Text}. Bitte gib einen Kategorienamen an, um den Text treffend zu kategorisieren!"*). Gut formulierte und ausreichend spezifizierte Prompts sind ein Schlüsselfaktor für zuverlässige Ergebnisse (Fischer et al., 2023). Um unerwünschte Prompt-Injektionen zu vermeiden - die auftreten können, wenn die Texte selbst Aufforderungen enthalten, welche die Wahrscheinlichkeit des weiteren Dialogverlaufs in eine unerwünschte Richtung verschieben - kann es hilfreich sein, die Anweisungen an das Ende des Prompts zu stellen. Optional kann die Vorverarbeitung auch eine agentenbasierte oder Retrieval-basierte Prompt-Ergänzung (*Retrieval Augmented Generation*, Xu et al., 2021; Fischer et al., 2024b) beinhalten, um zusätzliche Informationen dynamisch einzubinden (z.B. können für jeden Text semantisch ähnliche Texte, Kategorien und/oder Fakten abgerufen und dem Prompt hinzugefügt werden, über eine Formulierung wie *"Bitte beachte die folgenden Informationen aus unserer Datenbank: {Information}"*). Generell müssen Prompt-Templates entsprechend den Anforderungen des KI-Modells

formuliert werden (was die meisten KI-Systeme heutzutage automatisch tun können), und die Implikationen der vorherigen Schritte auf die Vorverarbeitung sollten berücksichtigt werden:

  - Bei zustandslosen Anfragen ist die Vorverarbeitung relativ einfach. Je nach Korpus - und unabhängig von der gewählten Verarbeitungsart - ist es in der Regel empfehlenswert, bestimmte Zeichen oder Zeichensequenzen aus der Verarbeitungseinheit zu entfernen oder zu ersetzen (z. B. „*\n\n*“ oder „**“, sowie andere Zeichensequenzen, die für Ihren Arbeitsablauf oder das Prompt-Template relevant sein können);
  - Bei zustandshaltigen Anfragen muss auch der aktuelle Stand der Analyse berücksichtigt werden, sofern vorhanden (z. B. *“Bitte gib einen Kategorienamen an, um den folgenden Text entsprechend zu kategorisieren: {Text}; Die folgende Liste von Kategorien - die zum aktuellen Text passen können oder auch nicht - ist aus früheren Texten bekannt: {Zustand}“*);
  - Bei der Massenverarbeitung umfasst die Vorverarbeitung in der Regel die Formatierung der einzelnen Verarbeitungseinheit, um Kodierungseinheiten, Kontexteinheiten und/oder Kategorien eindeutig zu klären (z. B. *“Gegeben seien die folgenden Texte (jeweils durch Semikolon getrennt): {Semikolon-getrennte Liste von Texten}; Ordne jeden dieser Texte mindestens einer der folgenden Kategorien zu: {nummerierte Liste von Kategorien}. Gib dazu die Antworten unverändert wieder und füge in Klammern die Nummer/n der entsprechenden Kategorie/n hinzu!”*).

- Bei der Verarbeitung wird der Prompt an ein KI-System gesendet, um eine Antwort zu generieren. Dies ist der zentrale Teil der KI-basierten Inhaltsanalyse, und es wird dringend empfohlen, den spezifizierten Arbeitsablauf anhand einer kleinen Anzahl von Verarbeitungseinheiten zu testen (z. B. auf der Grundlage fiktiver Daten oder einer kleinen, zufällig aus dem Korpus ausgewählten Menge), bevor der gesamte Korpus analysiert wird. Besonderes Augenmerk sollte auf den Datenschutz und andere Implikationen der geltenden rechtlichen Anforderungen (z. B. die KI-Verordnung der Europäischen Union) gelegt werden, insbesondere wenn personenbezogene oder sensible Daten im Prompt nicht *on-premises* oder in kontrollierbarer Weise verarbeitet werden.

- Bei der Nachverarbeitung werden aus den bisher generierten Antworten nützliche Informationen extrahiert. Je nach dem im Vorverarbeitungsschritt formulierten Prompt können dies Zahlen (z.B. Nummern der zugewiesenen Kategorien) oder Texte (z.B. formulierte Kategorien oder Paraphrasen) sein. Es wird dringend empfohlen, den Extraktionsprozess intensiv zu testen, um zu überprüfen, ob er wie erwartet funktioniert. Treten Fehler auf, so können diese manchmal durch eine Änderung der Vorverarbeitung (z.B. durch Prompt-Engineering) oder der Nachverarbeitung behoben werden (z.B. durch eine Vereinfachung/Kürzung der Antwort – man denke beispielsweise an das Entfernen von Textteilen nach Zeilenumbrüchen, das Ersetzen bestimmter Zeichen wie Einrückungen, Sternchen, Semikolons, o.ä.). Im Falle einer zustandshaltigen Verarbeitung beinhaltet dies auch die Formulierung eines neuen

Zustands (z. B. eine nummerierte Liste von Kategorien, die bei der Vorverarbeitung der nächsten Verarbeitungseinheit verwendet werden).

- Die Nachverarbeitung nach der Verarbeitung der letzten Verarbeitungseinheit spielt eine besondere Rolle. Hier kann es sinnvoll sein, die gesammelten Ergebnisse aller vorangegangenen Prozesse einzubeziehen, um die Ergebnisse zu verdichten oder zu vereinheitlichen (z.B. durch Clustering der Texteinbettungen aller generierten Kategorienbezeichnungen). Diese Art der Nachbearbeitung kann auch eine zusätzliche inhaltliche Analyse der erzeugten Ergebnisse umfassen (z.B. um hierarchische Strukturen von Kategorien oder fallübergreifende Zusammenfassungen auf der Grundlage fallbezogener Zusammenfassungen zu erhalten, vgl. Döring & Bortz, 2016, 559 - für ein Beispiel siehe den Abschnitt über KI-basierte Zusammenfassungen unten);
- Die Nachverarbeitung kann auch eine interaktive Aufgabe sein, bei der man als Analytiker*in den Output der KI zu mehreren Zeitpunkten während des Workflows verarbeitet. Ein Sonderfall der Nachverarbeitung liegt vor, wenn jeder Verarbeitungsschritt nicht automatisch, sondern interaktiv von Analyst*innen verarbeitet wird *(„Cyborg-Modus“)*. In diesem Fall können Analyst*innen nach jedem Verarbeitungsschritt nützliche Informationen extrahieren und bei Bedarf zusätzliche Informationen - oder mehrere Vorschläge für jede Verarbeitungseinheit - anfordern (in einem mehrschrittigen Dialog). Für diesen Fall empfiehlt es sich jedoch, jeden Dialog detailliert zu dokumentieren und im Allgemeinen für jede Verarbeitungseinheit einen neuen Dialog/Chat zu beginnen (um übermäßig komplexe Abhängigkeiten bei der Generierung von Antworten zu vermeiden).

5. Implementierung

Der fünfte Schritt ist die eigentliche Umsetzung (und Überwachung bzw. das Monitoring) des festgelegten Arbeitsablaufs. Für jede Verarbeitungseinheit wird empfohlen, Informationen über den Prozess in einer Datei zu protokollieren und sicherzustellen, dass jeder Prozess der Analyse gut dokumentiert ist und potenziell (a) inspiziert und (b) nach jeder verarbeiteten Einheit fortgesetzt werden kann (z. B. im Falle von Verbindungsfehlern) - im Hinblick auf eine LLM-basierte Analyse ist es in der Regel sinnvoll, Informationen über Prompt(s), Antwort(en), die extrahierten Informationen und/oder den jeweiligen Zustand der Analyse für jede Verarbeitungseinheit zu protokollieren. Bei zustandshaltigen Anfragen ist es von besonderer Bedeutung, den Zustand der Analyse nach jeder verarbeiteten Einheit zu überwachen (da Fehler in diesem Fall die folgenden Verarbeitungseinheiten beeinflussen). Generell ist es bei der KI-basierten Inhaltsanalyse empfehlenswert, die Verarbeitung genau zu überwachen und bei Bedarf zum Beginn (oder zu einem anderen der bereits absolvierten Schritte) des Ablaufmodells zurückzukehren. Selbst wenn keine Fehler auftreten, besteht grundsätzlich die Option, die Analyse (oder Varianten der Analyse) mehrfach (oder mit mehreren Modellen) zu wiederholen, um die Zuverlässigkeit und Robustheit der Ergebnisse zu bewerten.

6. Evaluation

Der sechste Schritt ist die Auswertung und Berichterstattung der Ergebnisse. Im Hinblick auf eine LLM-basierte Analyse können die aus den KI-basierten Antworten extrahierten Informationen im Hinblick auf die zu Beginn festgelegte Forschungsfrage interpretiert und möglichst reproduzierbar berichtet werden. Mit Blick auf die Validität der Analyse wird empfohlen, alle KI-generierten Antworten zu überprüfen. Vor dem Hintergrund beschränkter Ressourcen empfiehlt es sich, eine Stichprobe der KI-generierten Antworten auszuwerten und auf auffällige Ausreißer zu prüfen (z.B. ist es oft sinnvoll, neben einer Zufallsstichprobe von Antworten auch insb. die längsten und kürzesten Antworten auf ihre Validität zu prüfen).

5. Beispiele KI-basierter Inhaltsanalysen

Im folgenden Kapitel werden verschiedene Ansätze zur KI-basierten Inhaltsanalyse, die in verschiedenen Projekten umgesetzt wurden, und die dabei gewonnenen Erkenntnisse näher erläutert. Der Fokus liegt auch hier explizit auf LLM-basierten Analysen.

Zu beachten ist, dass die bereitgestellten Prompt-Templates in deutscher Sprache formuliert und auf Korpora deutscher Texte angewendet wurden. Zum Zeitpunkt der Analysen, die überwiegend im Jahr 2025 durchgeführt wurden, kamen in den meisten Fällen KI-Modelle aus der Gemma-Serie von Google und deren Finetunings zum Einsatz (z.B. unsloth/gemma-3-12b-it-GGUF oder mradermacher/gemma-2-27b-it-SimPO-37K-GGUF) mit System-Prompts wie *„Du bist ein renommierter Sozialwissenschaftler“*.

Im Folgenden werden unsere Erfahrungen mit verschiedenen Arten von KI-basierten Analysen getrennt nach Kategorien (Zusammenfassung, Explikation, Induktion und Deduktion) dargestellt, um einen Einblick in die jeweils entwickelten Prompts und die Art der erzielten Ergebnisse zu geben. Der Schwerpunkt liegt nicht auf den Ergebnissen der Analysen (die ganz überwiegend Arbeitsstände darstellen und zu gegebener Zeit in separaten Artikeln und Berichten veröffentlicht werden – der jeweils aktuelle Stand kann beim Autor erfragt werden). An dieser Stelle geht es darum, exemplarisch aufzuzeigen wie verschiedene Arten von KI-basierten Inhaltsanalysen prinzipiell realisiert werden können und was dabei jeweils zu beachten ist.

5.1. KI-basierte Zusammenfassung

Die Zusammenfassung ist eine der Kernkompetenzen moderner KI und eine tragende Säule in der KI-gestützten QIA. Bei dieser Art der Inhaltsanalyse nimmt die KI Verarbeitungseinheiten entgegen und generiert auf dieser Basis Schritt für Schritt eine verdichtete Zusammenfassung des Korpus, meist durch Reduktion redundanter Aussagen und durch Bündelung verwandter Aussagen.

Hinweise für die Praxis

Ein prototypisches Beispiel für einen Prompt zur KI-basierten Zusammenfassung von Texten wäre folgendes (mit Platzhaltern für Text und Forschungsfrage in geschweiften Klammern):

Gegeben sei folgender Text: „{Text}“
Paraphrasiere alle angesprochenen Erkenntnisse zur Frage „{Forschungsfrage}“!

Im Projekt transform.by haben wir zum Beispiel Interviews zum Wandel der Automobilindustrie in Deutschland geführt und unter anderem KI-basierte Zusammenfassungen zur Generierung von Paraphrasen eingesetzt.

Zunächst führten wir eine Analyse durch, bei der wir jedes Interviewtranskript als Verarbeitungseinheit verwendeten, wobei wir das folgende Prompt-Template anwandten:

> *„Gegeben sei das folgende Transkript eines Interviews:\n\n\"{Transkript}\"\n\nBitte paraphrasiere in einem Satz alle Erkenntnisse, die in diesem Interview zur Transformation der Automobilindustrie angesprochen werden (Trends und Treiber, Chancen und Herausforderungen, Organisations- und Personalentwicklung, etc.)!"*

Die generierten Paraphrasen wurden dann mittels Massenverarbeitung konkateniert und weiter verdichtet über folgendes Prompt-Template:

> *„Gegeben seien die folgenden Paraphrasen von Interviews:\n\n\"{Paraphrasen}\"\n\nBitte erstelle eine stichpunktartige Zusammenfassung aller angesprochenen Punkte. Gliedere deine Darstellung nach Punkten, die (a) in allen Interviews angesprochen wurden, (b) in mehreren aber nicht allen Interviews angesprochen wurden, (c) nur in einzelnen Interviews angesprochen wurden."*

Das Ergebnis war eine strukturierte Liste mit allen wichtigen Punkten aus den Interviews in Form einer groben Zusammenfassung. Für eine detaillierte Zusammenfassung führten wir darüber hinaus noch eine zweite Analyse durch, bei der wir die einzelnen Passagen jedes Interviews als Verarbeitungseinheiten verwendeten. Für jede Passage eines jeden Interview-Transkripts haben wir das folgende Prompt-Template verwendet:

> *"Gegeben sei die folgende Passage aus einem Interview:\n\n\"{Passage}\"\n\nBitte paraphrasiere in einem Satz alle Erkenntnisse, die in dieser Passage zur Transformation der Automobilindustrie angesprochen werden (Trends und Treiber, Chancen und Herausforderungen, Organisations- und Personalentwicklung, etc.)! "*

Die Paraphrasen wurden dann hierarchisch zusammengefasst - zunächst interviewspezifisch und dann interviewübergreifend, um die verfügbaren Informationen schrittweise zu verdichten. Diese zweite Analyse war zwar deutlich komplexer, lieferte aber auch deutlich mehr Informationen auf unterschiedlichen Detailebenen durch die Umschreibung auf unterschiedlichen Hierarchieebenen.

Im Rahmen desselben Projekts führten wir auch eine KI-gestützte Inhaltsanalyse von Stellenanzeigen durch, wobei unser Ziel darin bestand, eine große Anzahl von Stellenanzeigen im Hinblick auf die gesuchten Kompetenzen zusammenzufassen. Angesichts der großen Zahl zu analysierender Texte beschlossen wir, jede Stellenanzeige einzeln zu verarbeiten. Für jede Stellenanzeige verwendeten wir das folgende Prompt-Template, um die Informationen, an denen wir interessiert waren, zu verdichten:

> *"Gegeben sei folgende Stellenanzeige:\n\n\"{Text}\"\n\nDeine Aufgabe ist es, aufzulisten, welche Kompetenzen (im Sinne von Kenntnissen, Fähigkeiten, Fertigkeiten, oder Einstellungen) in der*

Stellenanzeige explizit gesucht werden. Formuliere jeden Eintrag knapp als ein bis maximal drei Worte in deutscher Sprache, beschränkte dich auf explizit gesuchte Kompetenzen, verzichte auf Klammern oder Kommentare, und trenne die Kompetenzen jeweils durch durch Strichpunkte (;)!"

Das Ergebnis war eine durch Strichpunkte getrennte Liste von Kompetenzen für jede Stellenanzeige. Interessanterweise kann es bei einigen KI-Systemen vorkommen, dass sie die Anweisungen ignorieren und bestimmte Formulierungen in den Texten ernst nehmen, wenn der zu analysierende Text am Ende des Prompt-Templates steht (z. B. generierte das KI-System in einem Fall eine Bewerbung auf eine Stellenanzeige im Bereich Datenwissenschaft, anstatt die Fähigkeiten, Fertigkeiten und Kompetenzen wie gefordert zu extrahieren).

5.2. KI-basierte Explikation

Bei der Explikation besteht das Ziel der Analyse darin, *„zu einzelnen fraglichen Textteilen (Begriffen, Sätzen usw.) zusätzliches Material hinzuzufügen, das das Verständnis erweitert, das die Textstelle expliziert, erklärt und ausdeutet"* (Mayring, 2022, 66). Wie Mayring (2022) betont, kann die explizierende qualitative Inhaltsanalyse von der einfachen Übersetzung über die Explikation des direkten Textzusammenhangs (enge Kontextanalyse) bis hin zur Berücksichtigung externer Materialien oder freier Assoziationen der Analyst*innen (breite Kontextanalyse) reichen. Im Fokus steht grundsätzlich eine Kontextualisierung des Explikandums (bemerkenswert sind hier Parallelen zum Konzept der *Retrieval Augmented Generation* im KI-Kontext, vgl. Fischer et al., 2023).

Hinweise für die Praxis

Ein prototypisches Beispiel für einen Prompt zur KI-basierten Explikation von Texten wäre folgendes (mit Platzhaltern für Text und Forschungsfrage in geschweiften Klammern):

Gegeben sei folgender Text: „{Text}"
Erläutere, inwiefern der Text Hinweise gibt auf die Frage „{Forschungsfrage}"!

Im Hinblick auf KI-basierte Varianten dieser Form der Inhaltsanalyse können textgenerierende LLMs neben Zusammenfassungen (s.o.) auch Übersetzungen, Interpretationen, Begründungen oder Erklärungen für einzelne Analyseeinheiten eines Korpus generieren.

In einem Projekt untersuchten wir zum Beispiel Stellenanzeigen und wollten wissen, in welchem Kontext das Thema Studienabbruch angesprochen wurde. Wir suchten daher nach relevanten Stellenanzeigen und wandten KI-basierte Explikation an, indem wir jede Stellenanzeige separat verarbeiteten und nach einer Explikation des Kontexts fragten, in dem das Thema Studienabbruch erwähnt wurde. Für jede Stellenanzeige verwendeten wir folgendes Prompt-Template, um den Kontext der Erwähnung des Studienabbruchs zu erläutern:

„Gegeben sei die folgende Stellenanzeige:\n\n\"{Text}\"\n\nDeine Aufgabe ist es, den Kontext zu erläutern, in dem das Thema \"Studienabbruch\" im Text angesprochen wird (z. B. direkte

*Ansprache von Studienabbrecher*innen, Erwähnung formaler Aspekte wie ECTS-Punkte oder indirekte Ansprache von Studienabbrechern über Formulierungen wie \"nicht-traditionelle Karriere\"). Paraphrasiere jede Passage der Stellenanzeige, die in direktem Zusammenhang mit dem Thema steht, in jeweils einem Satz!"*

Mit Blick auf die Explikation auf Basis externer Materialien (breite Kontextanalyse, siehe Mayring, 2022) eignen sich hier insbesondere KI-Systeme, die Zugriff auf eine Datenbank mit relevanten Dokumenten oder auf die Ergebnisse einer relevanten Websuche haben (je nach LLM und Datenquelle(n) können diese Ergebnisse manuell kuratiert oder dem LLM zur Laufzeit automatisch per *Retrieval Augmented Generation* zur Verfügung gestellt werden). Eng verwandt mit der deduktiven Kategorienzuweisung (s.u.) können auch verschiedene Arten von explikativen Informationen automatisch dem Material zugeordnet werden, etwa im Sinne einer Ad-hoc-Klassifikation/Schlüsselwortbildung durch ein textgenerierendes LLM (Single-Shot-Klassifikation oder Multi-Shot-Klassifikation), im Sinne eines Topic Labeling (Rieger et al., 2024) oder durch Annotation semantisch verwandter Materialien aus einem Korpus von Klassifikationskriterien, Kategorien, Schlagworten o.ä. (zur automatisierten Verschlagwortung siehe Fischer et al., 2023; zur automatisierten Kategorisierung siehe Fischer & Dörpinghaus, 2024).

5.3. KI-basierte Induktion

Induktive Kategorienbildung, der Kern beim Entdecken emergenter Themen und Muster in qualitativen Daten, profitiert erheblich von den Fähigkeiten moderner KI. Bei dieser Art der induktiven Inhaltsanalyse nimmt die KI Verarbeitungseinheiten entgegen und generiert das, was im Hinblick auf theoretische Konzepte (Kategorien) folgt.

Hinweise für die Praxis

Ein prototypisches Beispiel für einen Prompt zur KI-basierten Induktion von Kategorien wäre folgendes (mit Platzhaltern für Text und Forschungsfrage in geschweiften Klammern):

Gegeben sei folgender Text: „{Text}"
Gehe bei der Auswertung in zwei Schritten vor: (1) Erstelle zunächst eine Liste aller angesprochenen Kategorien und definiere für jede Kategorie, wie sie von den übrigen Kategorien abzugrenzen ist (\n"Kategoriensystem\n"). (2) Liste für jede Kategorie alle angesprochenen Punkte auf, die sich der Kategorie eindeutig zuordnen lassen und gib jeweils die Nennungshäufigkeit an (\"Kategorienzuordnung\"). (3) Interpretiere die Ergebnisse mit Blick auf die Frage „{Forschungsfrage}"!

Im Projekt KIPerWeb (Fischer, Lorenz, Pabst, 2024; Fischer et al., 2023; Fischer et al., 2022) haben wir beispielsweise eine Online-Umfrage zur Personalisierung der beruflichen Weiterbildung durchgeführt. Die Umfrage enthielt eine Reihe offener Fragen zu verschiedenen Oberkategorien von Beeinträchtigungen (psychisch/seelisch, körperlich, sensorisch & geistig).

Wir wendeten KI-basierte Induktion von Kategorien an, um ein detailliertes Kategoriensystem auf der Grundlage der Freitextantworten unserer Teilnehmenden zu erhalten. Für jede Oberkategorie wurden alle Freitextantworten aneinandergereiht, als durch Semikolon getrennte Liste eingefügt und an eine KI geschickt mit der Bitte, (a) alle genannten Punkte aufzulisten (mit der jeweiligen Nennungshäufigkeit), (b) die Punkte in verschiedenen Kategorien von Teilnahmehindernissen zusammenzufassen und (c) schließlich die identifizierten Teilnahmehindernisse prägnant in einem Satz zusammenzufassen. Für jede Oberkategorie von Beeinträchtigungen haben wir das folgende Prompt-Template verwendet:

> *„Auf die Frage \"Haben Sie bestimmte Beeinträchtigungen, die sich auch auf das Lernen oder das\nAufsuchen von Unterrichtsräumen auswirken können? (Mehrfachantworten sind möglich)\" wurden in der Kategorie \"{Oberkategorie}\" folgende Antworten formuliert (jeweils durch Strichpunkte getrennt):\n\n"{Antworten}"\n\nBitte erstelle eine Liste aller genannten Punkte (jeweils mit Nennungshäufigkeit), gliedere die Punkte in unterschiedliche Kategorien von Teilhabebarrieren, und fasse abschließend die ermittelten Teilhabebarrieren in einem Satz prägnant zusammen!"*

Bemerkenswert ist, dass dieser Ansatz der oben beschriebenen KI-basierten Zusammenfassung ähnelt, aber darauf abzielt, Kategorien aus den Rohdaten zu induzieren (ähnlich dem traditionellen Topic Modeling, vgl. Bittermann et al., 2018). Die Antworten der KI wurden manuell überprüft und zu einem kohärenten Kategoriensystem von Teilhabebarrieren mit Kategorien und Unterkategorien zusammengefasst.

Dieser Ansatz funktionierte im Allgemeinen recht gut, und wir hatten nur geringfügige Korrekturen hinsichtlich der logischen Kohärenz des generierten Kategoriensystems vorzunehmen. Es sollte jedoch betont werden, dass LLMs - die normalerweise darauf trainiert sind, plausible Texte oder Dialogvervollständigungen zu generieren - nicht unbedingt fehlerfrei zählen oder den Regeln der Mathematik folgen (insbesondere ältere und/oder kleinere Modelle): In Bezug auf die Häufigkeit der Erwähnungen haben wir konkret festgestellt, dass die Zahlen nicht immer ganz korrekt waren, obwohl sie als heuristischer Indikator für die Relevanz immer noch nützlich waren (weil sie die Größenordnungen angemessen repräsentierten, um wichtige Kategorien von weniger wichtigen zu unterscheiden).

Einen ähnlichen Ansatz der KI-basierten Induktion haben wir auch auf das Ergebnis der KI-basierten Zusammenfassungen der oben genannten Stellenanzeigen angewandt: Ausgehend von einer Liste von Kompetenzen - die eine große Anzahl von synonymen Formulierungen für identische Konstrukte enthielt - haben wir die KI-basierte Induktion mittels Massenverarbeitung (*Bulk Processing*) angewandt, um Kategorien von Kompetenzen zu induzieren. Aufgrund der großen Anzahl von Formulierungen war sowohl die Aggregation aller Begriffe als auch der paarweise Vergleich aller Begriffe auf Synonymität nicht praktikabel. Stattdessen wendeten wir eine komplexere Art der Nachbearbeitung an: Zunächst bestimmten wir die 100 häufigsten Begriffe und wendeten eine KI-basierte Induktion zur Identifizierung von Oberkategorien auf diese Begriffe an (i.S.v. *„Gruppiere diese Kategorien zu Oberkategorien, indem du synonyme oder eng verwandte Kategorien als JSON-Liste kombinieren (z. B. \"[{,Problem*

Solving':[,Problem Solving','Problem Solving']]\")."). Anschließend wendeten wir dann eine iterative KI-basierte Deduktion an, um jeden Eintrag aus der vollständigen Liste aller Begriffe einer der Kategorien zuzuordnen ("*Gegeben sei folgendes Kategoriensystem:\n\n{Kategorien}\n\nDeine Aufgabe ist es, die folgende Formulierung einer passenden Kategorie zuzuordnen:\n\n\„{Eintrag}\"\n\nAntworte direkt mit dem Code der Kategorie, die am besten passt. Wenn keine Kategorie passt, antworten Sie mit 999 für \"Andere\". Erläutere anschließend deine Antwort Antwort!*"). Auf diese Weise konnten wir auf Basis verbreiteter Begriffe ein Kategoriensystem entwickeln, das auf alle Materialien in unserem Korpus anwendbar war.

Wie das letztgenannte Beispiel zeigt, ist eine Massenverarbeitung nicht immer umsetzbar: Wenn die Liste der zu kategorisierenden Texte für das KI-System zu groß ist, kann es vorkommen, dass die Massenverarbeitung aus technischen Gründen verworfen werden muss (z.B. im Falle eines KI-Modells mit zu kleinem Kontextfenster oder wenn die Analyse im Falle einer Massenverarbeitung zu viel Zeit oder Ressourcen in Anspruch nehmen würde), auch wenn sie aus methodischer Sicht wünschenswert sein mag. In diesem Fall kann es sinnvoll sein, zunächst eine Teilmenge der wichtigsten Texte über Massenverarbeitung zusammenzufassen (z.B. die Top-k-Aussagen aus der Liste der Aussagen, geordnet nach ihrer Häufigkeit der Erwähnung oder anderen Relevanzindikatoren). Zusätzliche Texte könnten dynamisch über eine zustandsabhängige Verarbeitung hinzugefügt werden (z. B. durch Hinzufügen einer Formulierung wie *„Bitte ordnen Sie den Text einem Kategorie-Label aus der Liste zu, oder geben Sie gegebenenfalls ein neues Kategorie-Label an"* oder über einen Prompt, der die aktuelle Liste der Kategorien angibt, und eine Aktualisierung der Kategorien-Liste nach jeder Verarbeitungseinheit).

5.4. KI-basierte Deduktion

Bei der deduktiven Kategorienanwendung wird die KI eingesetzt, um bereits vorhandene Kategorien auf die Materialien des Korpus anzuwenden bzw. die Materialien entsprechend zuzuordnen. Allgemein geht es bei der deduktiven Inhaltsanalyse mit KI darum, dass die KI auf Basis einer Reihe von Prämissen (z. B. theoriebasierte Kategorien und/oder Definitionen von Konzepten) Schlussfolgerungen für eine bestimmte Verarbeitungseinheit generiert. Bei einer großen Anzahl von Kategorien können Methoden wie Retrieval-Augmented Generation angewandt werden, um Kategorienzuordnungen auf der Grundlage einer Teilmenge von Kategorien zu generieren (Fischer et al., 2023).

Hinweise für die Praxis

Ein prototypisches Beispiel für einen Prompt zur KI-basierten Deduktion zu Kategorien wäre folgendes (mit Platzhaltern für Text und Kategoriensystem in geschweiften Klammern):

Gegeben sei folgender Text: „{Text}"
Ordne diesen Text einer der folgenden Kategorien zu: „{Kategoriensystem}".
Gib die ID der gewählten Kategorie an („ID") und begründe deine Antwort („Begründung")!

Natürlich gibt es mehrere Varianten bzw. Formen deduktiver Kategorienanwendung, mit jeweils unterschiedlichen Vorzügen, die man KI-basiert umsetzen kann: Die der Analyse zugrunde zu legenden *„Strukturierungsdimensionen"* der deduktiven/strukturierenden Inhaltsanalyse sind nach Mayring (2016, S.99) entweder inhaltlich (um Material zu bestimmten Themen oder Inhaltsbereichen zu extrahieren und zusammenzufassen), formal (um die innere Struktur des Materials nach bestimmten formalen Strukturierungsaspekten herauszufiltern), typisierend (um einzelne markante Merkmale im Material auf einer Typisierungsdimension zu finden und zu beschreiben) oder skalierend (um Merkmale für einzelne Dimensionen in Form von Skalenpunkten zu definieren und das Material entsprechend zu bewerten).

In einem Projekt haben wir beispielsweise eine große Anzahl anonymisierter *„Leistungs- und Verhaltensbeurteilungen"* (LuV) mittels skalierender deduktiver Kategorienanwendung ausgewertet, um Hinweise auf die Wirksamkeit einer Maßnahme auf aggregierter Ebene zu erhalten. Für jede LuV wurde das folgende Prompt-Template verwendet (vgl. Bauer et al., 2025):

> *"Die Berufseinstiegsbegleitung verfolgt das Ziel, die Chancen der Schülerinnen und Schüler auf einen erfolgreichen Übergang in eine berufliche Ausbildung deutlich zu verbessern und diese jungen Menschen zu stabilisieren. Zum Abschluss einer Berufseinstiegsbegleitung wurde für eine*n Teilnehmende*n folgende Leistungs- und Verhaltensbeurteilung (LuV) erstellt, die im Folgenden ausgewertet werden soll.\nBitte fasse die LuV in einem Satz zusammen ('Zusammenfassung') und bewerte in Klammern dahinter in Prozent die aktuelle Ausprägung umfassender Handlungskompetenz, verstanden als Bereitschaft und Befähigung des Einzelnen, sich in fachbezogenen, gesellschaftlichen und privaten Situationen sachgerecht durchdacht sowie individuell und sozial verantwortlich zu verhalten ('Handlungskompetenz'). Begründe deine Bewertung abschließend kurz in einem Satz ('Bewertungsgrundlage').\n\nLuV:\n\n{LuV}"*

Das Ergebnis war eine Schätzung des Kompetenzniveaus für jede LuV (die auf einer aggregierten Ebene ausgewertet wurden, um die LuV zu Beginn und am Ende der Maßnahme zu vergleichen), gefolgt von einer kurzen Erläuterung der Bewertung (vgl. Bauer et al., 2025). Zu beachten ist, dass wir in diesem Fall den Text an das Ende des Prompts gesetzt haben, da er lang war und rein deskriptive Informationen aus vertrauenswürdigen Quellen enthielt (d.h. ohne das Risiko von Prompt-Injektionen). Die Bitte um eine Begründung der Bewertung führte unseres Erachtens zu plausibleren Einschätzungen der Kompetenz. Wir haben auch verschiedene Arten der Bewertung ausprobiert (z. B. Schulnoten), fanden diese aber weniger geeignet und die Varianz zwischen den LuVs war in diesem Fall gering.

In einem anderen Projekt hatten wir eine Liste von Tätigkeiten und Merkmalen, die von Fachleuten als wichtig für einen Beruf angesehen wurden. Unser Ziel war es, jeden Eintrag in dieser Liste der Kompetenz zuzuordnen, mit der er am engsten verbunden ist. Daher wendeten wir für jeden Eintrag in der Liste das folgende Prompt-Template an, um ihm eine Kategorie zuzuordnen:

"Zentrales Ziel der Berufsschule ist die Förderung der Entwicklung einer umfassenden Handlungskompetenz. Ein Tätigkeitsscreening hat ergeben, dass im Arbeitsbereich ‚{Bereich}' folgendes besonders wichtig ist: \n\n ‚{Text}'.\n\nDeine Aufgabe ist es, diese Formulierung einer Kompetenz zuzuordnen. Gehe Schritt für Schritt vor: Überlege zunächst, welche Kompetenzen in diesem Kontext eine Rolle spielen und welche nicht (\„Ableitung\"). Dann gib den Code der Kompetenz ein, die in diesem Kontext am relevantesten ist (\„Kategoriecode \")."

In diesem Fall baten wir das KI-Modell konkret, *„Schritt für Schritt"* vorzugehen, und fragten nach einer Ableitung einer Lösung, bevor eine Kategorie ausgewählt werden sollte, anstatt im Anschluss an die Kategorienzuordnung nach einer Erklärung der Kategorienzuordnung zu fragen. Wie erwartet, verbesserte sich dadurch die Anzahl der plausiblen Kategorienzuordnungen weiter - tatsächlich erwiesen sich in diesem Fall bei der manuellen Überprüfung 99 % der Kategorienzuordnungen als plausibel und korrekt.

6. Diskussion

Wir sind überzeugt, dass KI bereits heute zur Unterstützung und teilweisen Automatisierung der qualitativen Inhaltsanalyse eingesetzt werden kann. Kontinuierliche menschliche Kontrolle und Verfeinerung bleiben jedoch während des gesamten Prozesses bis heute unerlässlich, um Genauigkeit, Tiefe und letztendlich Aussagekraft der Erkenntnisse zu gewährleisten. Derzeit nutzen immer mehr Menschen KI-Systeme wie ChatGPT und ähnliche Methoden zur Zusammenfassung und Analyse von Texten, was auch durchaus bisweilen kritikwürdig ist (Mühlhoff & Henningsen, 2024). Im Hinblick auf die KI-basierte Inhaltsanalyse ist insofern zu betonen, dass die Qualität und Vertrauenswürdigkeit der Ergebnisse nicht nur von der Verwendung guter Werkzeuge, sondern auch von deren kompetenter Benutzung abhängt.

In diesem Beitrag haben wir das Ablaufmodell von Fischer (im Druck) vorgestellt, welches die wesentlichen Schritte einer KI-basierten Inhaltsanalyse beschreibt und so ein systematisches Vorgehen ermöglicht. Die Verwendung von kostenlosen Open-Source- bzw. Open-Weights-Modellen kann unseres Erachtens ein weiterer wichtiger Baustein für wissenschaftliche Anwendungen der KI-basierten Inhaltsanalyse sein, da dadurch gewährleistet wird, (a) dass die Analysen bestmöglich reproduziert werden können, (b) dass die Datenverarbeitung vor Ort und datenschutzkonform erfolgt und (c) dass das Modell langfristig, unverändert und kosten-los zur Verfügung steht.

Wie die oben ausgeführten Beispiele zeigen, erwies sich die Übertragung traditioneller Techniken der Inhaltsanalyse in das Zeitalter der KI als vielversprechend:

Die traditionelle Methode zur induktiven Kategorienbildung (siehe Mayring, 2022) lässt sich als eine Reihe von zustandshaltigen Anfragen zur Textgenerierung an menschliche Textgeneratoren beschreiben (d. h., neue Kategorien werden der Liste der für die folgenden Verarbeitungseinheiten berücksichtigten Kategorien hinzugefügt). Im Gegensatz zu menschlichen Analysierenden ist KI jedoch auch in der Lage, große Textmengen als Ganzes zu analysieren (i.S.v. Massenverarbeitung) oder jede Textpassage selektiv so zu verarbeiten, dass kein Wissen über vorherige oder zukünftige Textpassagen einfließt (i.S.v. zustandslosen Anfragen).

Deduktive Inhaltsanalyse hingegen wurde traditionell eher als eine Reihe von isolierten zustandslosen Abfragen durch menschliche Zuweisungsgeneratoren implementiert. KI-basierte Deduktion kann diesen Prozess nach unserem Dafürhalten weitgehend automatisieren und sogar bereichern. Beispielsweise kann KI für jede Verarbeitungseinheit nicht nur die Einheit einer Kategorie zuordnen, sondern gleichzeitig auch eine Begründung für jede Zuweisung generieren (was die Zuweisung selbst in der Regel noch zuverlässiger und leichter überprüfbar macht).

Auch in Hinblick auf zusammenfassende und explizierende Inhaltsanalysen ermöglichen große Sprachmodelle bereits heute die Erstellung hervorragender Textentwürfe für Zusammenfassungen und Erläuterungen. Diese können die Analyse großer Textmengen erheblich erleichtern und sowohl einen Überblick (über Zusammenfassungen) als auch Einblicke (über Erläuterungen) in große Textmengen in deutlich kürzerer Zeit als bisher ermöglichen.

Insgesamt zeigen die Beispiele in diesem Artikel hoffentlich, dass die KI-basierte Inhaltsanalyse viele Möglichkeiten bietet, Teile der qualitativen Inhaltsanalyse zu unterstützen und zu automatisieren. Als Analytiker*innen sollten wir dieses Potenzial nicht ungenutzt lassen, wobei natürlich der Mensch stets im Mittelpunkt einer professionellen und ethisch vertretbaren Arbeit mit KI stehen muss (Fischer et al., 2023, 32).

7. Literatur

Ahrbeck, B., Lehmann, R., Fickler-Stang, U., Kretschmer, A., & Maué, E. (2009). *ENEBS. Evaluationsstudie Berliner Schülerfirmen.* Berlin: Humboldt-Universität zu Berlin.

Bauer, P., Fischer, A., Herdegen, N., Kugler, P, Boockmann, B., & Hecker, K. (2025). Evaluation der Berufseinstiegsbegleitung in Sachsen (BerEbS) - Auswirkungen der landesspezifischen Änderungen - Abschlussbericht. *f-bb-Bericht 01/2025.*

Berelson, B. (1952). *Content analysis in communication research.* Free Press.

Berding, F., Tykhonova, Y., Pargmann, J., Leube, A., Riebenbauer, E., Rebmann K., & Slopinski, A. (2025). *aifeducation: Artificial Intelligence for Education.* https://github.com/cran/aifeducation

Bergmann, D., Richter, k.E., Fischer, A.,& Peters, M. (2024). Digitale Lehr- und Lernsettings in pflegeberuflichen Weiterbildungen – eine webbasierte Analyse von Weiterbildungsangeboten. In: U. Weyland, W. Koschel, K. Reiber, L. Dorin, M. Peters (Hrsg.), *Digitalisierung in den Gesundheitsberufen (pp. 178-196).* Leverkusen: Budrich.

Bittermann, A., & Fischer, A. (2018). How to Identify Hot Topics in Psychology Using Topic Modeling. *Zeitschrift für Psychologie, 226(1).* https://doi.org/10.1027/2151-2604/a000318

Dell'Acqua, F., McFowland III, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., ... & Lakhani, K. R. (2023). Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality. *Harvard Business School Technology & Operations Mgt. Unit Working Paper*, (24-013).

Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation (Vol. 5)*. Heidelberg: Springer.

Feuerriegel et al. (2025). Using natural language processing to analyse text data in behavioural science. *Nature Reviews Psychology, 4,* 96-111. https://doi.org/10.1038/s44159-024-00392-z

Fischer, A. (im Druck). The Future of Qualitative Content Analysis in the Age of Artificial Intelligence in M. Tiemann & J. Dörpinghaus (Eds.), *Advances in Computational Social Sciences*.

Fischer, A. (2023) Toot 111296359168891768 on Chaos.Social. Technical Report. Available online: https://chaos.social/@AFischer1985/111296359168891768 (accessed on 1 January 2025).

Fischer, A., & Dörpinghaus, J. (2024). Web Mining of Online Resources for German Labor Market Research and Education: Finding the Ground Truth? *Knowledge, 4(1),* 51-67. https://doi.org/10.3390/knowledge4010003

Fischer, A., Jöchner, A., & Dauser, D. (2024). *Open Educational Resources (OER) und Künstliche Intelligenz (KI) – Entwicklungschancen für die berufliche Weiterbildung.* f-bb-online 03/24.

Fischer, A., Hilse, P., & Schütt-Sayed, S. (2023). Ausbildungsordnungen und Rahmenlehrpläne – Spiegel der Bedeutung nachhaltiger Entwicklung. In: I. Pfeiffer & H. Weber (Hrsg.), *Zum Konzept der Nachhaltigkeit in Arbeit, Beruf und Bildung–Stand in Forschung und Praxis* (pp. 281-302). Leverkusen: Budrich.

Fischer, A., Jöchner, A., Pabst, C., Lorenz, S., & Schley, T. (2023): KI-basierte Personalisierung berufsbezogener Weiterbildung. Ein Praxisleitfaden für Bildungsanbieter. In: S. Kretschmer & I. Pfeiffer (Hrsg.), *Leitfaden für die Bildungspraxis 73 (3-39).* https://doi.org/10.3278/9783763976447

Fischer, A., Lorenz, S., & Pabst, C. (2024): Empfehlungen zur beruflichen Weiterbildung. Entwicklung eines KI-basierten Entscheidungsmanagements. In: *Berufsbildung in Wissenschaft und Praxis (BWP), 1/24, 32-34.*

Fischer, A., Pabst, C., Jöchner, A., Lorenz, S., & Schley, T. (2022). Möglichkeiten, Bedarfe und Wünsche bezüglich der Personalisierung berufsbezogener Weiterbildung. *bwp@, 43,* 1-24.

Fischer, A., & Pfeiffer, I. (2019). Zugänge zur beruflichen Aufstiegsfortbildung nach Bachelorabschluss? Eine Dokumentenanalyse von Fortbildungsordnungen. *Berufsbildung in Wissenschaft und Praxis (BWP)*, 5/19, 56-57.

Friese, S. "Qualitative data analysis with ATLAS. ti." (2019): 1-344.

Friese, S. (2024): Künstliche Intelligenz in der qualitativen Forschung: ein neuer Horizont. Keynote at the 10th Swith Methods Festical. St. Gallen: Switzerland.

Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The journal of socio-economics, 40(1),* 35-42.

Glaser, B., & Strauss, A. (1967). *Discovery of grounded theory: Strategies for qualitative research*. Routledge.

Hofmann, V., Kalluri, P. R., Jurafsky, D. & King, S. (2024) AI generates covertly racist decisions about people based on their dialect. *Nature, 633, 147–154.*

Krippendorff, K. (1969). Models of messages: three prototypes. In: G. Gerbner, O.R. Holsti, K. Krippendorff, G.J. Paisly & Ph.J. Stone (Eds.), *The analysis of communication content.* New York: Wiley.

Kuckartz, U., & Rädiker, S. (2023). Qualitative content analysis: Methods, practice and software. Sage.

Kuckartz, U., & Rädiker, S. (2019). *Analyzing qualitative data with MAXQDA*. Cham: Springer International Publishing.

Kuckartz, U., Ebert, T., Rädiker, & S., Stefer, C. (2009). *Evaluation online – Internetgestützte Befragung in der Praxis*. Heidelberg: Springer.

Lämmel, U., & Cleve, J. (2008). *Künstliche Intelligenz (Vol. 3).* München: Hanser.

Mayring, P. (1985). Qualitative Inhaltsanalyse in: Jüttemann, Gerd (Hrsg.): *Qualitative Forschung in der Psychologie.*

Marz, N., Warren, J. (2016). *Big Data – Entwicklung und Programmierung von Systemen für große Datenbmengen und Einsatz der Lambda-Architektur.* Frechen: mitp.

Mayring, P. (2000). Qualitative Content Analysis. *Forum: Qualitatice Social Resarch, 1(2).*

Mayring, P. (2016). *Qualitative inhaltsanalyse (12., überarbeitete Auflage).* Weinheim: Beltz.

Mayring, P. (2022). *Qualitative inhaltsanalyse (13., überarbeitete Auflage).* Weinheim: Beltz.

Mühlhoff, R. & Henningsen, M. (2024). *Chatbots im Schulunterricht: Wir testen das Fobizz-Tool zur automatischen Bewertung von Hausaufgaben.* https://arxiv.org/abs/2412.06651 (Dateiversion v5/2024-12-17, abgerufen am 07.01.2025)

Palomino, A., Fischer, A., Kuzilek, J., Nitsch, J., Pinkwart, N., & Paaßen, B. (2024). EdTec-QBuilder: A Semantic Retrieval Tool for Assembling Vocational Training Exams in German Language. *In: Kai-Wei Changi; Annie Lee; Nazneen Rajani (Hrsg.). Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: System Demonstrations. Annual Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL-2024), System Demonstrations, located at System Demonstrations Track, Mexico City, Mexico City, Mexico, ACL, 6/2024.* *10.18653/v1/2024.naacl-demo.3*

Schlögl, A., Hofer, N., & Böhme, R. (2023). Causes and effects of unanticipated numerical deviations in neural network inference frameworks. *Advances in Neural Information Processing Systems, 36.*

Schlögl, A., Kupek, T., & Böhme, R. (2021, June). iNNformant: Boundary samples as telltale watermarks. *Proceedings of the 2021 ACM Workshop on Information Hiding and Multimedia Security (pp. 81-86).*

Rieger, J., Peters, F., Fischer, A., Lauer, T., & Bittermann, A. (2024). *topiclabels: Automated Topic Labeling with Language Models*. https://cran.r-project.org/web/packages/topiclabels/index.html

Weizenbaum, J. (1977). *Computer Power and Human Reason. From Judgment to Calculation.* W H Freeman & Co.

Xu, Z., Cruz, M.J., Guevara, M., Wang, T., Deshpande, M., Wang, X., & Li, Z. (2024). *Retrieval-Augmented Generation with Knowledge Graphs for Customer Service Question Answering. In: Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 2905-2909).*

Außerdem zuletzt vom f-bb veröffentlicht

Bauer, P., Wittig, W., & Weber, H. (2024): *Stärkung der Ausbildungsbereitschaft von Betrieben: Wie der Transfer von Bildungsinnovationen gelingen kann. Arbeitshilfe für die Transferpraxis.* f-bb-online 02/24. https://www.f-bb.de/unsere-arbeit/publikationen/staerkung-der-ausbildungsbereitschaft-von-betrieben-wie-der-transfer-von-bildungsinnovationen-geling/

Dauser, D. (2024): *Beschäftigte mit Open Educational Resources (OER) und Künstlicher Intelligenz (KI) gezielt fördern! Eine Praxishilfe für die betriebliche Personalentwicklung im Mittelstand.* f-bb-online 04/24. https://www.f-bb.de/de/unsere-arbeit/publikationen/beschaeftigte-mit-open-educational-resources-oer-und-kuenstlicher-intelligenz-ki-gezielt-foerdern/

Dauser, D., & Utomo, M. (2025): *KI-Chatbots Marke Eigenbau?! Whitepaper mit Hintergrundinformationen, Empfehlungen und Praxistipps*. f-bb-online 01/25. https://www.f-bb.de/unsere-arbeit/publikationen/ki-chatbots-marke-eigenbau-whitepaper-mit-hintergrundinformationen-empfehlungen-und-praxistipps/

Fischer, A., Jöchner, A., & Dauser, D. (2024). *Open Educational Resources (OER) und Künstliche Intelligenz (KI) – Entwicklungschancen für die berufliche Weiterbildung.* f-bb-online 03/24. https://www.f-bb.de/unsere-arbeit/publikationen/open-educational-resources-oer-und-kuenstliche-intelligenz-ki/

Fischer, A., Jöchner, A., Pabst, C., Lorenz, S., Schley, T. (2023): KI-basierte Personalisierung berufsbezogener Weiterbildung. Ein Praxisleitfaden für Bildungsanbieter. In: S. Kretschmer & I. Pfeiffer (Hrsg.), *Leitfaden für die Bildungspraxis 73 (3-39).* https://doi.org/10.3278/9783763976447

Pabst, C., Jöchner, A., Fischer, A., Lorenz, S., & Schley, T. (2023): Modularisierung berufsbezogener Weiterbildung. Ein Praxisleitfaden für Bildungsanbieter. In: S. Kretschmer & I. Pfeiffer (Hrsg.), *Leitfaden für die Bildungspraxis 74*. https://doi.org/10.3278/9783763976461

Richter, K., & Müller, J. (2023): Berufliche Weiterbildung im Kontext der digitalen Transformation. Digitale Methoden und Medienformate zur Gestaltung beruflicher Bildungsinhalte. f-bb-online 04/23. https://www.f-bb.de/unsere-arbeit/publikationen/berufliche-weiterbildung-im-kontext-der-digitalen-transformation-digitale-methoden-und-medienformat/