



Abbildung 1: Automatisierte Berichterstellung mit KI. Eigene Abbildung, erstellt mit ChatGPT-4o (OpenAI).

Automatisierte Berichterstellung mit Künstlicher Intelligenz

Eine Machbarkeitsstudie für die Applikation *socialweb*, basierend auf bestehenden Journaleinträgen der Sozialen Arbeit

Masterarbeit

Studiengang

Autor

Betreuer

Auftraggeberin

Lead-Experte

Co-Expertin

Datum

MAS Leadership in Innovation and Technology

Matthias Andenmatten

Beat Kiener

GLAUX GROUP

Prof. Bramwell Kaltenrieder

Prof. Dr. Mascha Kurpicz-Briki

29.07.2025 (überarbeitete Version)

Hinweis

Diese Version der Masterarbeit wurde nach der Verteidigung vom 27.03.2025 überarbeitet. Es wurden folgende inhaltliche Anpassungen vorgenommen:

- Korrektur: Konkretisierung des verwendeten Sprachmodells (*GPT-4o* statt allgemeiner Verweis auf *ChatGPT*).
- Ergänzung: Verweis auf fehlerhafte Textgenerierungen (Halluzinationen) in Kapitel 8.1.1 und entsprechende Aufführung im Anhang 16.1.
- Redaktionelle Präzisierungen einzelner Begriffe und Formulierungen zur besseren Nachvollziehbarkeit.

Die Anpassungen dienen der wissenschaftlichen Korrektheit und methodischen Transparenz. Sie basieren auf den Rückmeldungen der Prüfungskommission.

Inhaltsverzeichnis

1	Abstract	5
2	Management Summary	6
3	Einleitung	7
3.1	Ausgangslage und Motivation	7
3.2	Problemstellung	8
3.3	Zielsetzung und Forschungsfragen	8
3.4	Abgrenzung	9
3.5	Aufbau der Arbeit	12
4	Stand der Forschung Teil 1 – Technologische Perspektiven	13
4.1	Grundlagen der KI für die Textverarbeitung	13
4.1.1	Einleitung	13
4.1.2	Vokabulare	16
4.1.3	Bag-of-Words	17
4.1.4	Neuronale Netzwerke	17
4.1.5	Rekurrente neuronale Netzwerke	19
4.1.6	Transformatoren	20
4.1.7	Trainieren von Modellen	21
4.1.8	Retrieval Augmented Generation	24
4.2	Modelle in der praktischen Anwendung	25
4.2.1	Modellauswahl	25
4.2.2	Trainingsdaten und ihre Implikationen	27
4.2.3	Prompting	28
4.2.4	Qualität der Eingabedaten	31
4.2.5	Evaluierungsmethoden	32
4.2.6	Agent AI	34
4.2.7	Explainable AI	34
4.3	Implementierung	35
4.3.1	Hardwareanforderungen	35
4.3.2	Implementationsverfahren	35
4.3.3	Nachhaltigkeit	36
4.4	Zusammenfassung	36
4.4.1	Erkenntnisse zur Zielsetzung	37
4.4.2	Erkenntnisse zu den Forschungsfragen	37
5	Stand der Forschung Teil 2 – Perspektiven der Sozialen Arbeit	40
5.1	Anwendungsmöglichkeiten von KI in der Sozialen Arbeit	40
5.2	Verantwortung und Ethik in der Anwendung	40
5.3	Dokumentation und Berichtswesen	42
5.4	Zusammenfassung	43
5.4.1	Erkenntnisse zur Zielsetzung	44
5.4.2	Erkenntnisse zu den Forschungsfragen	44
6	Methoden Teil 1 – Technologische Umsetzung	46
6.1	Proof of Concept und Bestimmung der Stossrichtung	46
6.2	Datenbeschaffung, Analyse und Vorbereitung	48
6.3	Umsetzung der Stossrichtung	48
6.4	Prompting	48
6.5	Bias-Analyse und Reduktion	48
6.6	Evaluierung	48
7	Methoden Teil 2 – Wirtschaftliche Perspektiven	49
8	Technische Ergebnisse	50
8.1	Proof of Concept und Bestimmung der Stossrichtung	50

8.1.1 Stossrichtung 1: Interne Architektur	50
8.1.2 Stossrichtung 2: <i>SwissGPT</i> von <i>AlpineAI</i>	52
8.1.3 Stossrichtung 3: <i>ChatGPT</i> von <i>OpenAI</i>	52
8.1.4 Evaluation und Auswahl Stossrichtung	53
8.2 Datenbeschaffung, Bereinigung und Anonymisierung	53
8.2.1 Datenbeschaffung	53
8.2.2 Bereinigung und Anonymisierung	54
8.3 Datenanalyse	54
8.3.1 Quantitative Analyse	54
8.3.2 Qualitative Analyse	55
8.3.3 Zusammenfassung und Handlungsempfehlungen	57
8.4 Erstellung von Berichten	57
8.4.1 Einfaches Prompting	57
8.4.2 Ergänzendes und verbessertes Prompting	58
8.4.3 Bias-Analyse und Reduktion	61
8.4.4 Evaluierung der Ergebnisqualität	61
8.5 Zusammenfassung und Fazit	62
8.5.1 Erkenntnisse zur Zielsetzung	63
9 Wirtschaftliche Ergebnisse	64
10 Diskussion	65
10.1 Einfluss der Datenqualität auf MLOps	65
10.2 Akteurinnen und Akteure im MLOps-Prozess	67
10.2.1 GLAUX GROUP	68
10.2.2 Institutionen	70
10.2.3 Externe Akteurinnen und Akteure	71
10.2.4 Zusammenfassung und Fazit	73
10.3 Ethische Herausforderungen und Verantwortung	74
10.4 Reflexion der methodischen Vorgehensweise	75
10.5 Alternative KI-Ansätze in der Praxis	77
10.5.1 Handbuch-Chatbot	77
10.5.2 Automatisierte Arbeitsplanung	78
11 Schlussfolgerungen, Empfehlungen und Fazit	79
11.1 Beantwortung der Hauptforschungsfrage	79
11.2 Beantwortung der technischen Forschungsfragen	80
11.3 Beantwortung der Forschungsfragen im Zusammenhang mit der Sozialen Arbeit	82
11.4 Beantwortung der wirtschaftlichen Forschungsfragen	84
11.5 Handlungsempfehlungen für die Praxis	85
11.6 Ausblick	86
12 Tabellenverzeichnis	89
13 Abbildungsverzeichnis	90
14 Glossar	92
15 Literaturverzeichnis	95
16 Anhang	99
16.1 Halluzination des Modells « <i>google/pegasus-multi_news</i> »	99
16.2 Quantitative Datenanalyse	99
16.3 Qualitative Datenanalyse von Dossier K-2021-00003	104
16.4 Qualitative Datenanalyse von Dossier K-2021-00060	105
16.5 Qualitative Datenanalyse von Dossier K-2021-00079	107
17 Selbständigkeitserklärung und Nutzung von KI	109

1 Abstract

Die automatisierte Berichterstellung mithilfe von Künstlicher Intelligenz (KI) gewinnt zunehmend an Bedeutung. Diese Arbeit untersucht, ob und inwiefern KI-Modelle wie Large Language Models (LLMs) zur automatisierten Erstellung von Berichten in der Sozialen Arbeit eingesetzt werden können. Diese umfassen strukturierte schriftliche Dokumentationen zu Fallverläufen, Interventionen und Entscheidungen, die als Grundlage für Fachpersonen, Institutionen und Behörden dienen. Als Fallbeispiel dient die Softwarelösung *socialweb*, in der Berichte auf Basis bestehender Journaleinträge generiert werden sollen.

Die Analyse zeigt, dass KI-basierte Berichterstellung technisch realisierbar, jedoch mit erheblichen Einschränkungen verbunden ist. Während sprachliche Kohärenz und Textqualität der automatisch erzeugten Inhalte hoch sind, bestehen zentrale Herausforderungen hinsichtlich Datenqualität, Transparenz und Reproduzierbarkeit. Halluzinationen, fehlende Konsistenz und inhaltliche Unschärfen erschweren den direkten Einsatz in der Praxis. Zudem fehlen klare methodische Standards zur strukturierten Erfassung und Validierung der Daten, was die Qualität der generierten Texte erheblich beeinflusst.

Die Ergebnisse verdeutlichen, dass eine vollautomatische Berichterstellung in der Sozialen Arbeit derzeit nicht umsetzbar ist. Stattdessen bietet KI Potenzial als unterstützendes Werkzeug, etwa durch automatisierte Prüfung von Inhalten, Strukturierungshilfen oder Vorschläge zur Berichtgenerierung, die von Fachpersonen geprüft und angepasst werden. Um dieses Potenzial zu nutzen, müssen Institutionen verbindliche Dokumentationsstandards etablieren und Softwareanbieterinnen und -anbieter gezielte technische Anpassungen vornehmen, um die Qualität und Verlässlichkeit von KI-generierten Berichten zu verbessern.

Die Arbeit leistet einen Beitrag zur Fachdiskussion über KI in der Sozialen Arbeit und zeigt auf, welche technischen, methodischen und ethischen Rahmenbedingungen für einen erfolgreichen KI-Einsatz geschaffen werden müssen. Zudem enthält sie konkrete Handlungsempfehlungen für Institutionen sowie für Unternehmen aus der Softwareentwicklung, um eine sinnvolle Integration von KI in bestehende Arbeitsprozesse zu ermöglichen.

2 Management Summary

Ausgangslage

Die Digitalisierung in der Sozialen Arbeit schreitet voran und der Einsatz von Künstlicher Intelligenz (KI) zur automatisierten Berichterstellung wird zunehmend diskutiert. In der Praxis sind Fachpersonen mit einem hohen Dokumentationsaufwand konfrontiert, der viel Zeit in Anspruch nimmt. Die Softwarelösung *socialweb* könnte KI nutzen, um Berichte auf Basis bestehender Journaleinträge zu generieren. Diese Masterarbeit untersucht die technischen, methodischen und organisatorischen Herausforderungen einer solchen Implementierung.

Zentrale Erkenntnisse

Die Analyse zeigt, dass KI-basierte Berichterstellung grundsätzlich möglich, aber mit erheblichen Einschränkungen und Risiken verbunden ist:

- Datenqualität als kritischer Faktor: Die Konsistenz und Strukturierung der Eingabedaten beeinflussen die Qualität der generierten Berichte erheblich. Fehlende Standards in der Datenerfassung führen zu ungenauen oder unvollständigen KI-gestützten Berichten.
- Halluzinationen und mangelnde Transparenz: Large Language Models (LLMs) wie GPT können nicht verifizierbare oder fehlerhafte Inhalte generieren. Da diese Berichte schwer nachvollziehbar sind, bleibt eine fachliche Überprüfung durch Menschen unverzichtbar.
- Datenschutz und Anonymisierung: Die Nutzung externer KI-Modelle erfordert strenge Datenschutzmassnahmen, insbesondere eine vollständige Anonymisierung sensibler Daten oder die Verwendung lokaler KI-Modelle.
- Wirtschaftliche und ethische Herausforderungen: Der Einsatz von KI muss nicht nur technisch, sondern auch ethisch vertretbar sein. Die Verantwortung und Haftung für fehlerhafte Berichte liegt weiterhin bei den Fachpersonen und Institutionen.

Handlungsempfehlungen

Eine direkte vollautomatische Integration von KI in die Berichterstellung ist derzeit nicht sinnvoll. Stattdessen bietet KI Potenzial als unterstützendes Werkzeug, das Fachpersonen bei der Strukturierung, Formulierung und Validierung von Berichten zu entlasten. Um dieses Potenzial zu nutzen, werden folgende Massnahmen empfohlen:

Für die *GLAUX GROUP*

- Technische Infrastruktur optimieren: Erweiterung von *socialweb* zur besseren Strukturierung und Verarbeitung von Daten für zukünftige KI-Anwendungen.
- KI-gestützte Prüfmechanismen einführen: Automatisierte Erkennung von Inkonsistenzen und Fehlern bei der Dateneingabe.
- Weitere PoCs und Pilotprojekte initiieren: Entwicklung zusätzlicher Proofs of Concept (PoCs) sowie von Prototypen, um die Praxistauglichkeit der KI-gestützten Berichterstellung schrittweise und insbesondere in realen Anwendungsszenarien zu testen.

Für Institutionen

- Datenqualität verbessern: Einführung verbindlicher Dokumentationsstandards und konsequente Nutzung bereits vorhandener *socialweb*-Funktionen.
- Interdisziplinäre Zusammenarbeit fördern: Enge Abstimmung zwischen Sozialer Arbeit, IT, Fachverbänden und Hochschulen zur Entwicklung praxisnaher KI-Standards.
- Schulungen für Fachpersonen: Sensibilisierung für den Umgang mit KI-gestützten Assistenzsystemen zur Qualitätssicherung der Berichte.

Fazit und Ausblick

KI kann Fachpersonen in der Sozialen Arbeit bei der Dokumentation unterstützen, aber aktuell nicht ersetzen. Die Implementierung erfordert eine enge interdisziplinäre Zusammenarbeit sowie die Schaffung klarer technischer, methodischer und rechtlicher Rahmenbedingungen. Langfristig könnte eine Kombination aus strukturierten Daten, verbesserten KI-Modellen und standardisierten Prozessen die Qualität und Effizienz der Berichterstellung in der Sozialen Arbeit steigern. Die nächsten Schritte umfassen die Optimierung der Datenbasis, gezielte Pilotprojekte und eine kontinuierliche Weiterentwicklung der KI-Integration in *socialweb*.

3 Einleitung

3.1 Ausgangslage und Motivation

Der Autor dieser Arbeit verfügt über eine doppelte fachliche Basis: Er ist sowohl Informatiker mit Schwerpunkt Applikationsentwicklung als auch Sozialpädagoge. Diese Kombination ermöglichte es ihm, technologische Lösungen gezielt auf die Bedürfnisse der Sozialen Arbeit auszurichten. Ab 2008 entwickelte er die ersten beiden Versionen von *socialweb*, mit dem Ziel, Fachpersonen zu unterstützen und Prozesse in der Sozialen Arbeit zu optimieren. Gemeinsam mit seinem damaligen Geschäftspartner baute er das Start-up zu einer etablierten Grösse in der Sozialen Arbeit in der Deutschschweiz aus. Im Jahr 2022 wurde das Unternehmen in die *GLAUX GROUP AG* integriert, wodurch die Weiterentwicklung und Skalierung von *socialweb* in einem grösseren Unternehmenskontext sichergestellt wurde.

Die *GLAUX GROUP* ist ein wachsendes Softwareunternehmen, das massgeschneiderte Softwarelösungen für Behörden und verwandte Organisationen in der Schweiz entwickelt. Das Unternehmen koordiniert verschiedene operative Bereiche und sorgt dafür, dass die Anwendungen kunden- und aufgabenorientiert durch erfahrene Projektleiterinnen und Projektleiter eingeführt und betreut werden. Durch die Bündelung des umfangreichen Fachwissens in den verschiedenen Kompetenzbereichen ist das Unternehmen in der Lage, spezialisierte Lösungen anzubieten. Eigene Entwicklungsprojekte, aber auch Zukäufe von Firmen haben die Produktpalette in den letzten Jahren stetig erweitert. Aktuell bietet das Unternehmen Software für die digitale Verwaltung von Behörden, Softwarelösungen für die Bereiche Justiz, elektronische Signatur, Dokumentenverarbeitung und Outputmanagement, Klinikmarketing sowie die Standardsoftware *socialweb* für die Soziale Arbeit an.

Die Anwendung *socialweb* wurde in den letzten Jahren komplett neu entwickelt und basiert auf einem hochmodernen Technologie-Stack. Durch einen hohen Automatisierungsgrad ist die Aktualität der Software gewährleistet und es kann zeitnah auf neue Anforderungen des Marktes reagiert werden. In die Entwicklung der Anwendung sind das Fachwissen über die soziale bzw. sozialpädagogische Arbeit und die langjährige Erfahrung im Heimbereich eingeflossen. Die äusserst positiven Rückmeldungen der Institutionen sowie ein konstanter Zulauf zeugen davon, dass *socialweb* im Markt breit akzeptiert ist und die Branche prägt.

Socialweb wird in über 200 sozialen Einrichtungen eingesetzt. Die Anwendung ist modular aufgebaut und umfasst 20 Module. Um die fachlichen und administrativen Bedürfnisse der Institutionen bestmöglich abzudecken, wählen die Institutionen die für sie passenden Module aus und stellen sich ihr Softwarepaket zusammen.

Um den weiteren Erfolg der Software zu sichern, den Aufwand für die Fachpersonen zu reduzieren und insbesondere die Marktstellung weiter auszubauen, soll *socialweb* mit weiteren Funktionen ausgestattet werden. In diesem Zusammenhang wurde in den letzten Monaten mehrfach von Kundinnen und Kunden angeregt, die Implementierung von KI in die Software zu prüfen.

3.2 Problemstellung

In unterschiedlichen Modulen erfassen Fachpersonen täglich Informationen über die betreuten Klientinnen bzw. Klienten. Je nach Auftrag und Einrichtung werden umfassende Daten gesammelt. In regelmässigen Abständen finden Standortgespräche mit den betreuten Personen, Familien, zuweisenden Stellen und medizinischen Fachpersonen statt, welche anschliessend protokolliert werden. Damit die Gespräche zielgerichtet ablaufen, bedarf es einer fundierten Vorbereitung. Journaleinträge werden thematisch zusammengefasst und dienen als Leitfaden für Gespräch und Bericht. Das Lesen der Einträge und die Einarbeitung sind zeitaufwendig und erfordern pro Sitzung mehrere Stunden.

Je nach Periodizität der Gespräche wird eine grosse Datenmenge bearbeitet und die zeitlichen Anforderungen steigen. Es stellt sich die Frage, wie der Prozess effizienter gestaltet werden kann, ohne die Qualität und Aussagekraft zu beeinträchtigen. Durch den Einsatz von KI könnte die thematische Zusammenfassung der vorhandenen Journaleinträge automatisiert erfolgen. Dies würde nicht nur den zeitlichen Aufwand erheblich reduzieren, sondern idealerweise auch die Informationen für die Gespräche optimieren. Eine zentrale Herausforderung ist die datenschutzkonforme Integration von KI in *socialweb* sowie die Sicherstellung der Ergebnisqualität.

3.3 Zielsetzung und Forschungsfragen

Ziel dieser Arbeit ist, die nachstehenden Forschungsfragen gezielt und methodisch fundiert zu analysieren und zu beantworten:

Hauptforschungsfrage

- Welche technischen, methodischen und organisatorischen Voraussetzungen sind erforderlich, um KI in der Softwarelösung *socialweb* für die automatisierte Berichterstellung zu implementieren?

Technische Forschungsfragen

- Welche technischen Anforderungen müssen erfüllt sein, um eine KI-basierte Berichterstellung in *socialweb* zu implementieren und gleichzeitig die datenschutzrechtlichen Vorgaben zu wahren?
- Welche methodischen Anpassungen in der Datenerfassung und Datenaufbereitung werden benötigt?
- Ist eine Datenanonymisierung zwingend erforderlich, um KI datenschutzkonform in *socialweb* zu integrieren?
- Wie lässt sich die Ergebnisqualität der KI-generierten Berichte bewerten und kann diese durch Techniken wie Prompting optimiert werden?
- Wie entstehen Verzerrungen (Bias) und mit welchen Methoden lassen sich diese minimieren?

Forschungsfragen im Zusammenhang mit der Sozialen Arbeit

- Welche spezifischen Herausforderungen und Anforderungen ergeben sich in der Sozialen Arbeit bei der Dokumentation und der Erstellung von Berichten?
- Welche ethischen Anforderungen und Verantwortlichkeiten müssen KI-Systeme in der Sozialen Arbeit erfüllen, um professionellen und sozialen Ansprüchen gerecht zu werden?
- Wie und in welchen Bereichen wird KI bereits heute in der Sozialen Arbeit eingesetzt?

Wirtschaftliche Forschungsfragen

- Welche wirtschaftlichen Aufwände sind erforderlich, um KI in *socialweb* zu integrieren und nachhaltig zu betreiben?
- Welches Kosten-Nutzen-Verhältnis ergibt sich durch die Implementierung von KI zur automatisierten Berichterstellung im Vergleich zur manuellen Erstellung?
- Wie hoch ist der ökonomische und funktionale Nutzen für die Institutionen und welcher Mehrwert ergibt sich daraus für die *GLAUX GROUP*?

3.4 Abgrenzung

Das Themenfeld ist breit gefächert und überschneidet sich mit zahlreichen anderen Disziplinen. Aufgrund der Breite ist eine inhaltliche Fokussierung erforderlich. Daher werden einige Themen nur gestreift, während andere vollständig ausgeklammert werden. Die folgenden Bereiche werden angesprochen, jedoch nicht vertieft analysiert:

- **Technologischer Fortschritt und ethische Verantwortung**

Im *Geleitwort* der Studie *Wertschöpfung neu gedacht* verweisen die Autoren darauf, dass sich der Fortschritt nicht aufhalten lässt [1]. Diese Annahme wird in der vorliegenden Masterarbeit als Grundlage übernommen, ohne eine Reflexion darüber anzustossen, ob und wie der technologische Fortschritt, insbesondere im Zusammenhang mit KI, gestaltet werden sollte.

Gillhuber et al. thematisieren das Spannungsfeld zwischen wirtschaftlichen Anreizen und der gezielten Nutzung von Daten für Forschung und Entwicklung. In ihrem Kapitel *Datenethik zwischen gesellschaftlichem Anspruch und betrieblicher Praxis* formulieren sie dazu ethische Prinzipien für den Einsatz von KI [2, S. 93 ff.]. Altenburger beschreibt wiederum zentrale Herausforderungen im Umgang mit KI, insbesondere in Bezug auf Fairness, Rechenschaftspflicht und Transparenz. Er zeigt auf, dass KI-basierte Entscheidungsfindung bestehende Voreingenommenheiten verstärken und das Vertrauen in solche Systeme erheblich beeinträchtigen kann [3, S. 15]. Daraus lässt sich ableiten, dass ethische Fragestellungen nicht nur während der Produktentwicklung relevant sind, sondern über die gesamte Lebensdauer hinweg kontinuierlich kritisch hinterfragt und durch klare Auditverfahren überprüft werden müssen.

Obwohl einige ethische Aspekte in den folgenden Kapiteln behandelt werden, erfolgt nur eine oberflächliche Auseinandersetzung mit der Thematik. Für weiterführende Informationen zum Thema Ethik und KI wird auf die öffentlich zugänglichen Dokumente *Recommendation on the Ethics of Artificial Intelligence* der UNESCO [4] und *Ethically Aligned Design* der IEEE [5] sowie auf weitere einschlägige Fachliteratur verwiesen.

- **Schulung, Verantwortung und Prozessanpassung in Institutionen**

Da Resultate von KI-gesteuerten Funktionen automatisch generiert werden, ist eine kritische Reflexion der Ergebnisse unerlässlich. Besonders in der Sozialen Arbeit, einem Bereich mit heterogenem technischem Verständnis der Mitarbeitenden, ist eine umfassende Schulung für den verantwortungsvollen Einsatz und die Interpretation von KI-Ergebnissen notwendig.

Linnemann et al. beschreiben, dass KI als unterstützendes Werkzeug im Arbeitsalltag betrachtet werden kann, jedoch nicht als Ersatz für qualifiziertes Fachpersonal. Die Autoren betonen, wie wichtig es ist, dass KI-Systeme transparent aufzeigen, mit welcher Wahrscheinlichkeit die gelieferten Resultate zutreffend sind. Es muss sichergestellt werden, dass KI-Ergebnisse sinnvoll eingesetzt und korrekt interpretiert werden. Daher wird den Institutionen empfohlen, gezielte Schulungen für Fachkräfte anzubieten und die Interdisziplinarität der beteiligten Entwickler bzw. Entwicklerinnen zu fördern [6, S. 8].

Eine weiterführende Orientierung für Institutionen kann die *Data-Literacy-Charta* des Stifterverbands bieten [7]. Während der Masterarbeit wird das Thema jedoch nicht weiter vertieft. Der Vollständigkeit halber wird festgehalten, dass die *GLAUX GROUP* ihr Produktportfolio durch das Anbieten entsprechender Schulungen und Informationsveranstaltungen sinnvoll ergänzen könnte.

Weiter sei darauf hingewiesen, dass durch die Implementierung von KI in die fallführende Applikation bestehende Abläufe der Institution angepasst oder überarbeitet werden müssen. Es wird empfohlen, die Prozesse methodisch sowie fachlich zu prüfen und neu auszurichten. Die möglichen Auswirkungen auf den Alltag der Institutionen und die Erarbeitung von Handlungsanweisungen werden bei der Masterarbeit ausgeklammert.

- **Methoden und Algorithmen**

Im Kapitel 4 *Stand der Forschung Teil 1 – Technologische Perspektiven* werden Methoden, Funktionsweisen und Algorithmen thematisiert. Eine detaillierte technische Analyse dieser Technologien oder eine Vertiefung mathematischer und statistischer Grundlagen erfolgt jedoch nicht. Die Masterarbeit konzentriert sich auf das Zusammenfassen von Texten mithilfe von NLP-Transformatoren. Andere Informationen, wie beispielsweise Bilder, die theoretisch ebenfalls zur Dokumentation der täglichen Arbeit genutzt werden könnten, werden bewusst ausgeklammert und nicht untersucht.

- **Datenbasis**

Als Datenbasis dient eine Teilmenge von Daten, welche in der Applikation *socialweb* durch Fachpersonen im Modul *Journal* erfasst wurden. In der Masterarbeit wurden ausschliesslich Daten von einer Institution verwendet. Diese hat auch die Auswahl der Dossiers (Akten) vorgenommen und den Umfang bestimmt. Da die Daten nur von einer Institution stammen, ist die Repräsentativität sehr eingeschränkt und die Ergebnisse könnten daher spezifische Eigenheiten widerspiegeln. Nach Abschluss der Masterarbeit wird empfohlen, die beschriebenen Methoden auf einen grösseren Datenkorpus aus weiteren Institutionen und zusätzlichen Teilgebieten der Sozialen Arbeit anzuwenden und die Resultate erneut kritisch zu prüfen.

Eine detaillierte Analyse zur Datenqualität und den durchgeführten Vorverarbeitungsschritten wird in den nachfolgenden Kapiteln behandelt.

- **Hardwareanforderungen**

Die Implementierung und der Betrieb grosser Sprachmodelle erfordern eine leistungsfähige technische Infrastruktur. Aspekte wie Prozessorleistung, Speicherbedarf und Energieeffizienz spielen eine zentrale Rolle, insbesondere beim Training und der Anwendung von KI-Modellen. In dieser Arbeit werden Hardwareanforderungen zwar thematisiert, jedoch nicht im Detail analysiert. Fragen zur optimalen Hardwarekonfiguration, Skalierbarkeit und Kosteneffizienz bleiben ausserhalb des Fokus der Untersuchung.

- **Nachhaltigkeit**

Der hohe Rechenaufwand moderner KI-Modelle führt zu einem erheblichen Energieverbrauch, was ökologische und wirtschaftliche Fragen aufwirft. Während es zunehmend Forschung zu energieeffizienten Algorithmen, sparsamen Architekturen und nachhaltigen Rechenzentren gibt, bleibt das Thema Nachhaltigkeit in dieser Arbeit nur ein Randaspekt. Eine detaillierte Betrachtung der CO₂-Bilanz von KI-Systemen oder möglicher Massnahmen zur Reduktion des ökologischen Fussabdrucks wird nicht vorgenommen.

Die folgenden Themen liegen ausserhalb des Fokus dieser Arbeit, sind jedoch für eine spätere Produkteinführung essenziell:

- **Gesetzgebung**

Im *AI Act* des *Europäischen Parlaments* wurden Richtlinien für die Anwendung Künstlicher Intelligenz festgelegt [8]. Obwohl die Gültigkeit der Gesetzgebung derzeit nur für die EU besteht, wird davon ausgegangen, dass auch die Schweizer Gesetzgebung hierzu weitere Empfehlungen und Richtlinien aussprechen wird. Während der Masterarbeit erfolgt kein weiterer Exkurs in diese Thematik, da es den Rahmen der Arbeit sprengen würde.

- **Datenschutz**

Die Daten von betreuten Personen sind laut dem *Bundesgesetz über den Datenschutz* besonders schützenswert [9, Art. 5]. Wenn Daten im Rahmen dieser Arbeit beschafft und bearbeitet werden, geschieht das unter Einhaltung der ISO-Norm 27'001 [10], nach welcher die *GLAUX GROUP* zertifiziert ist. Aufgrund der aktuellen Zertifizierung wird davon ausgegangen, dass die Gesetzgebung bereits umfassend eingehalten wird.

Die hier erarbeiteten Lösungsvorschläge müssen sich entsprechend datenschutzkonform verhalten. Die Geschäftsleitung und der Chief Information Security Officer (CISO) sind über die Inhalte und Schwerpunkte dieser Masterarbeit informiert. Die Datenbeschaffung und die Verarbeitung der sensitiven Daten sind mit ausgewählten Institutionen geklärt und vertraglich geregelt.

Zusätzlich sei darauf hingewiesen, dass für die Verarbeitung schützenswerter Daten die ausdrückliche Einwilligung der betroffenen Personen erforderlich ist. In der Schweiz gehören personenbezogene Daten nicht den Institutionen, sondern den betroffenen Personen selbst. Daher ist eine explizite Zustimmung erforderlich, insbesondere wenn die Daten für Zwecke wie das Training von Modellen verwendet werden [9, Art. 6].

- **Manipulation von Ergebnissen**

Untersuchungen zeigen, dass KI-Systeme durch sogenannte *Adversarial Attacks* gezielt getäuscht und dadurch Ergebnisse entscheidend manipuliert werden können [11]. Vor der Implementierung von KI in *socialweb* wird empfohlen, Methoden zur Reduzierung dieser Problematik zu prüfen und gegebenenfalls in das Produkt zu integrieren. Diese Thematik wird jedoch nicht weiter in dieser Arbeit behandelt.

- **Alternative Ansätze zur KI-gestützten Berichterstellung**

Laut Ertel gibt es verschiedene Ansätze zur automatisierten Berichterstellung. Neben Large Language Models (LLMs) spielen regelbasierte Systeme, heuristische Modelle und hybride KI-Methoden eine Rolle. Während regelbasierte Systeme auf vordefinierten Regeln basieren, nutzen heuristische Modelle Schätzverfahren zur Problemlösung. Hybride KI-Modelle kombinieren verschiedene Ansätze, um Vorteile mehrerer Methoden zu nutzen [12, S. 6 ff.]. Obwohl diese Methoden in spezifischen Szenarien nützlich sein können, liegt der Schwerpunkt dieser Arbeit auf der technischen Machbarkeit von LLMs in der Applikation *socialweb*. Diese Entscheidung basiert darauf, dass diese Technologien durch Fortschritte in Skalierbarkeit und Kontextverarbeitung erhebliche Verbesserungen in der natürlichen Sprachverarbeitung und der automatisierten Texterstellung erzielt haben [12, S. 339 f.].

- **Empirische Validierung der Ergebnisse**

Die Arbeit basiert auf einer technischen Analyse und theoretischen Betrachtung der KI-gestützten Berichterstellung. Eine direkte Evaluation durch Fachpersonen aus der Sozialen Arbeit oder durch praktische Tests wurde nicht durchgeführt. Stattdessen stehen die Machbarkeit und methodischen Herausforderungen im Vordergrund. Die tatsächliche Praxistauglichkeit und Akzeptanz der Lösung wurde nicht empirisch überprüft, könnte aber in einem nächsten Schritt durch Nutzerstudien oder Pilotprojekte untersucht werden.

- **Wirtschaftliche Methoden**

Neben den hier beschriebenen Themen umfasst die vorliegende Arbeit auch eine eingehende Betrachtung der Wirtschaftlichkeit. Die verwendeten Vorgehensweisen werden im Kapitel 7 *Methoden Teil 2 – Wirtschaftliche Perspektiven* dargestellt. Eine tiefergehende Auseinandersetzung mit den wirtschaftlichen Ansätzen sowie eine Hinterfragung des Stands der Technik oder eine Suche nach weiteren, potenziell moderneren Ansätzen erfolgen jedoch nicht; die Verfahren werden lediglich angewendet.

3.5 Aufbau der Arbeit

Die vorliegende Arbeit ist in folgende Abschnitte gegliedert:

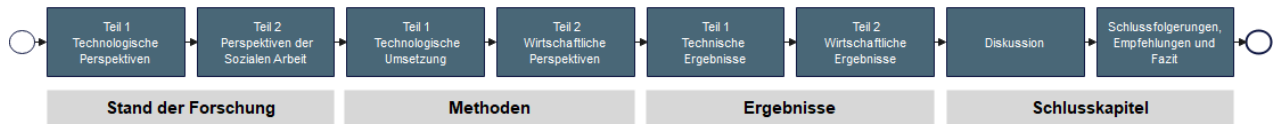


Abbildung 2: Aufbau der Masterarbeit. Eigene Abbildung.

Das Kapitel *Stand der Forschung Teil 1 – Technologische Perspektiven* untersucht die technischen Grundlagen und aktuellen Entwicklungen im Bereich der KI-gestützten Textzusammenfassung. Besonders das Grundlagenkapitel bildet eine unverzichtbare Basis für ein solides Verständnis und die technische Grundlage dieser Arbeit, weshalb darauf der grösste Fokus gelegt wird. Das Kapitel *Stand der Forschung Teil 2 – Perspektiven der Sozialen Arbeit* zeigt anschliessend die Sichtweise der Sozialen Arbeit auf.

Die Methodenkapitel *Methoden Teil 1 – Technologische Umsetzung* und *Methoden Teil 2 – Wirtschaftliche Perspektiven* wurden getrennt, um der technischen Methodik die nötige Tiefe zu geben, während bei der wirtschaftlichen Methodik der Fokus auf ökonomische Bewertung und Anwendbarkeit gelegt wird. Entsprechend werden auch die *Ergebniskapitel* getrennt geführt, um die unterschiedlichen Erkenntnisse spezifisch darzustellen und eine klare Analyse zu gewährleisten. Im Kapitel *10 Diskussion* werden die Fragestellungen und die Ergebnisse zusammengeführt, reflektiert und daraus entsprechende Erkenntnisse abgeleitet. Die Abrundung der Arbeit stellt das Kapitel *11 Schlussfolgerungen, Empfehlungen und Fazit* dar, bei welchem Handlungsempfehlungen für die Praxis abgeleitet und ein Ausblick auf zukünftige Forschungsansätze gegeben wird.

4 Stand der Forschung Teil 1 – Technologische Perspektiven

4.1 Grundlagen der KI für die Textverarbeitung

4.1.1 Einleitung

Dieses Kapitel gibt einen Überblick über die Entwicklung von KI. Der Fokus wird auf maschinelles Lernen und die Verarbeitung von Text gelegt, da diese Themen zentrale Aspekte der vorliegenden Arbeit darstellen.

Laut Heinlein und Huchler hat die Markteinführung von ChatGPT exemplarisch gezeigt, wie sich KI in den letzten Jahren fest in der Gesellschaft etabliert hat. Diese Technologie ist darauf ausgelegt, grosse, oft unstrukturierte Datenmengen zu analysieren, um Korrelationen und Muster zu erkennen [13, S. 15]. Dennoch ist KI kein Konzept des 21. Jahrhunderts, sondern wurde bereits 1956 von McCarthy und Minsky geprägt. Sie haben schon damals Technologien beschrieben, welche Maschinen ein intelligentes Verhalten ermöglicht haben [2, S. 150].

Ertel beschreibt die massgeblichen Technologien und Meilensteine in seinem Buch [12, S. 23]. Die in der untenstehenden Grafik dargestellte Zeitachse zeigt erste Meilensteine und für diese Arbeit wichtige Entwicklungen. Sie verdeutlicht, wie technologische Durchbrüche die Grundlage für moderne Anwendungen wie die Textverarbeitung geschaffen haben. Die Darstellung dient dazu, den Übergang von allgemeinen Fortschritten hin zu spezifischen Technologien wie dem maschinellen Lernen nachzuvollziehen. Einzelne Teile davon werden im weiteren Verlauf dieses Kapitels behandelt. Der Schwerpunkt liegt dabei auf der Verarbeitung von Sprache und Text.

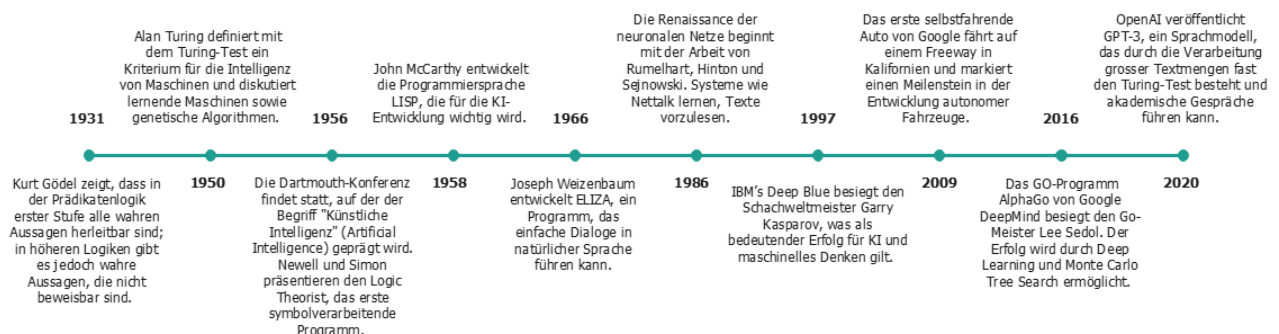


Abbildung 3: KI-Meilensteine in der Entwicklung [12, S. 23]. Eigene Abbildung.

Die chronologische Darstellung der Entwicklung von KI bietet die Grundlage, um spezifische Fortschritte wie maschinelles Lernen und Textverarbeitung detaillierter zu betrachten. In den folgenden Kapiteln wird aufgezeigt, wie diese Technologien die Art und Weise revolutioniert haben, wie Texte analysiert, interpretiert und generiert werden. Gleichzeitig werden verbleibende Herausforderungen benannt, um die Vielschichtigkeit der Thematik zu verdeutlichen.

Ein zentraler Aspekt bei der Einordnung von KI ist die Unterscheidung zwischen starker und schwacher KI. Schwache KI fokussiert sich auf klar definierte Anwendungsbereiche, wie etwa die Spracherkennung oder die Analyse von Texten. Sie dient als unterstützendes Werkzeug, das menschliches Denken simuliert, jedoch keine vollständige Autonomie besitzt. Starke KI hingegen beschreibt hypothetische Systeme, die über die intellektuelle Leistungsfähigkeit eines Menschen hinausgehen und autonom agieren könnten. Während schwache KI heute weit verbreitet ist, existiert starke KI bislang nur als theoretisches Konzept und wird kontrovers diskutiert, sowohl hinsichtlich ihrer technischen Umsetzbarkeit als auch ihrer ethischen Implikationen [13, S. 370].

Während sich diese Begrifflichkeiten theoretisch klar abgrenzen lassen, befindet sich die tatsächliche Entwicklung neuer KI-Technologien in einem ständigen Wandel. Um diesen Fortschritt besser einzuordnen, kann das Konzept des *Hype Cycles* herangezogen werden. Dieser beschreibt den typischen Entwicklungsverlauf neuer Technologien - von anfänglicher Begeisterung und überhöhten Erwartungen bis hin zu einer realistischen Einschätzung ihres tatsächlichen Potenzials. Die nachfolgende Grafik von Gartner aus dem Jahr 2024 zeigt verschiedene KI-Technologien und ordnet diese hinsichtlich ihrer erwarteten Reifezeit ein. Sie verdeutlicht, dass weiterhin hohe Erwartungen an KI-Technologien wie Generative KI (auch GenAI oder generative AI genannt) bestehen [14]. Im Zusammenhang mit der vorliegenden Arbeit wird erwartet, dass gerade GenAI entscheidende Potenziale mit sich bringt, die Technologie jedoch noch nicht vollständig ausgereift ist.

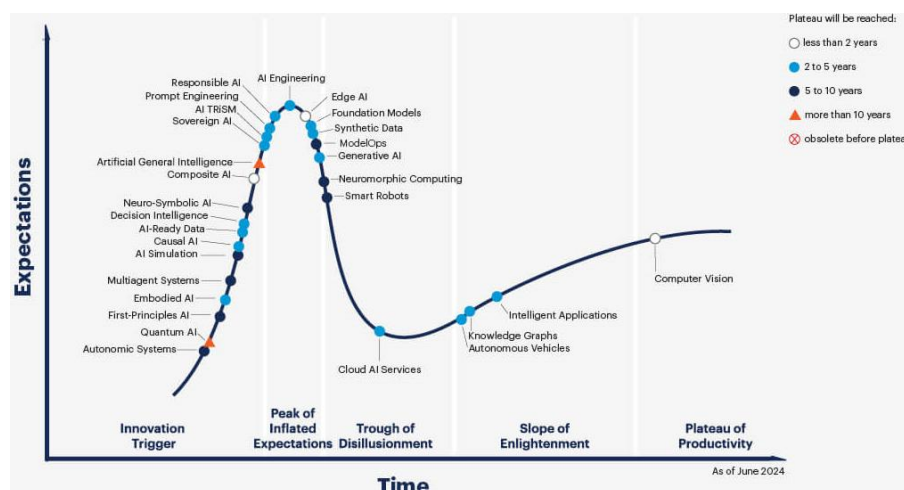
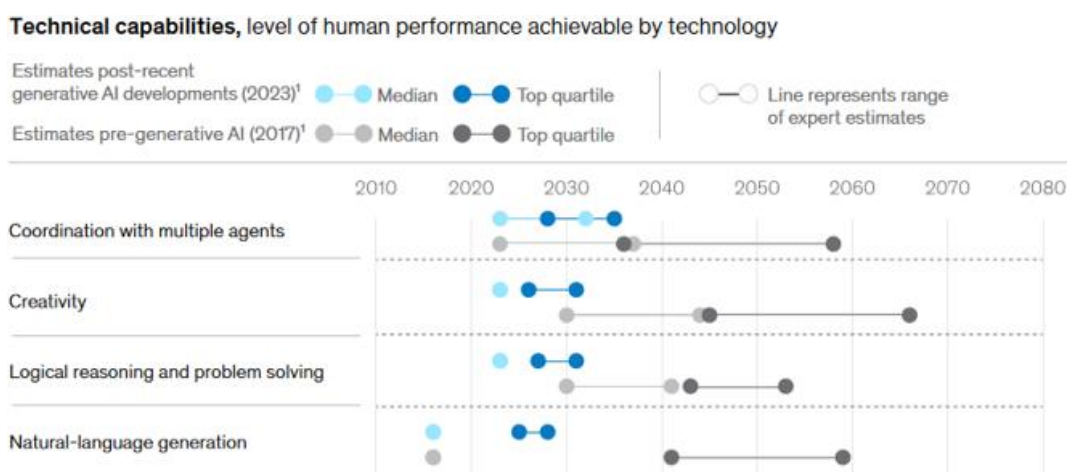


Abbildung 4: Hype Cycle für aufkommende KI-Technologien nach Gartner [14].

KI weckt derzeit hohe Erwartungen, und obwohl die Technologie noch nicht ganz ausgereift ist, macht die Studie von McKinsey Hoffnung, dass wir schon bald von Fortschritten profitieren können. Noch im Jahr 2017 ging man davon aus, dass die Innovation deutlich langsamer voranschreiten würde. Insbesondere generative KI hat jedoch in kurzer Zeit ein Leistungsniveau erreicht, das zuvor erst für spätere Zeitpunkte erwartet wurde. Die nachfolgende Grafik verdeutlicht, wie sich die technischen Fähigkeiten der KI in den kommenden Jahren entwickeln könnten. Sie zeigt, dass Technologien wie die Verarbeitung natürlicher Sprache und kreative Problemlösungen bereits grosse Fortschritte gemacht haben, auch wenn in vielen Bereichen noch Entwicklungspotenzial besteht [15]. Für die Soziale Arbeit eröffnen diese Entwicklungen spannende Möglichkeiten, etwa in der Analyse grosser Textmengen oder der Automatisierung von Berichterstellung und Dokumentation.



Technical capabilities, level of human performance achievable by technology

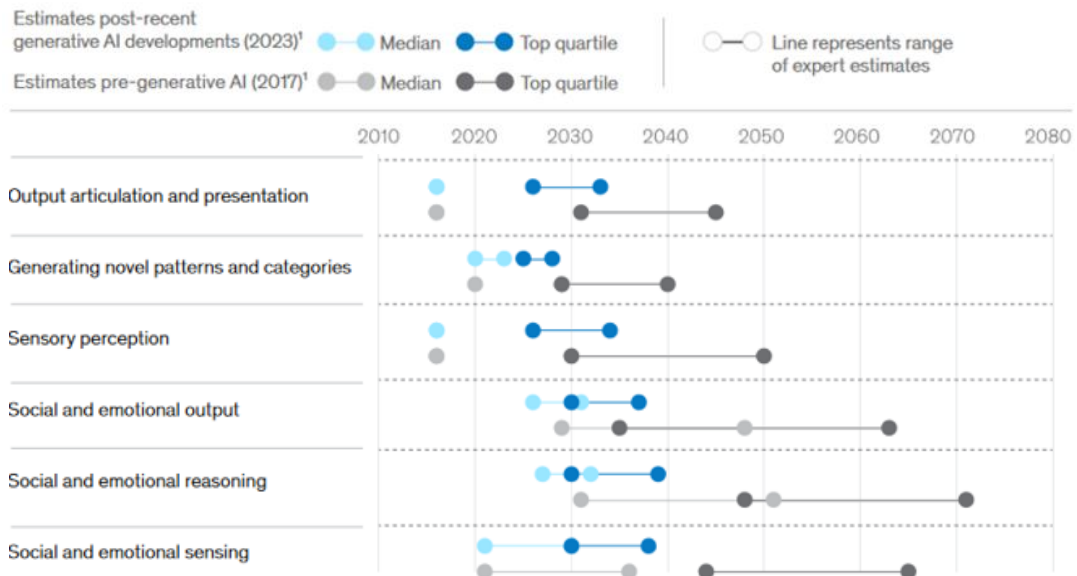


Abbildung 5: Menschliches Leistungsniveau, das durch Technologie erreicht werden kann [15].

Während die vorangegangene Abbildung einen Blick auf die zukünftigen technischen Möglichkeiten der KI wirft und zeigt, in welchen Bereichen noch Fortschritte notwendig sind, ist es ebenso wichtig, die aktuellen Teildisziplinen zu betrachten. Diese bilden die Grundlage für viele der heute bereits praxistauglichen Anwendungen und spielen eine entscheidende Rolle für die Themen dieser Arbeit. Die nachfolgende Abbildung bietet eine Einordnung dieser Teildisziplinen [1, S. 9].



Abbildung 6: Aktuelle Teildisziplinen der Künstlichen Intelligenz [1, S. 9].

Maschinelles Lernen (ML) und die Verarbeitung natürlicher Sprache (NLP) sind die zentralen Konzepte dieser Arbeit, da sie die Grundlage für moderne KI-Anwendungen bilden, insbesondere im Bereich der Textverarbeitung. ML ermöglicht es Computern, Muster in grossen Datenmengen zu erkennen und auf Basis dieser Erkenntnisse Vorhersagen zu treffen. NLP erweitert diese Fähigkeiten, indem es Maschinen erlaubt, menschliche Sprache zu analysieren, zu verstehen und neue Inhalte zu generieren. Beide Disziplinen sind essenziell, um KI-gestützte Werkzeuge für die Soziale Arbeit, wie die Automatisierung der Berichterstellung oder die Analyse umfangreicher Textdaten, zu entwickeln.

Nachfolgend werden die zentralen Technologien beschrieben, die diese Fortschritte ermöglichen. Dabei wird die Erstellung einfacher Vokabulare über Word Embeddings bis hin zu modernen Transformer-Architekturen nachvollzogen. Die Erklärungen und Visualisierungen in den folgenden Kapiteln lehnen sich stark an Alammr und Grootendorst an. Sie beschreiben grundlegende Konzepte des Maschinellen Lernens und der Verarbeitung natürlicher Sprache mithilfe anschaulicher Abbildungen.

4.1.2 Vokabulare

Die Grundlage des maschinellen Lernens für die Textverarbeitung beginnt mit der Erstellung eines Vokabulars, das die Basis für viele Sprachmodelle bildet. Ein Vokabular ist unerlässlich, da es den Rahmen definiert, innerhalb dessen Sprache analysiert und verarbeitet wird. Hierbei werden alle Token (kleinste sprachliche Einheiten) aus einem Text oder einer Sammlung von Texten, die als Korpus bezeichnet wird, systematisch erfasst. Je nach Modell können Token ganze Wörter, Wortteile oder sogar Zeichenfolgen sein. Ein Korpus kann aus Millionen von Wörtern bestehen, und durch den Prozess der Vokabularerstellung werden alle einzigartigen Token identifiziert. Die untenstehende Abbildung verdeutlicht diesen Prozess. Hier wird Text in kleinere Einheiten segmentiert und für die Verarbeitung durch Maschinen optimiert [16, Kap. Representing Language as a Bag-of-Words].

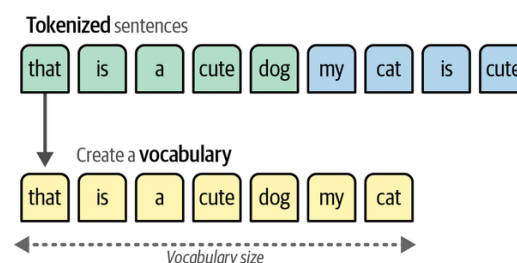


Abbildung 7: Erstellung eines Vokabulars [16, Kap. Representing Language as a Bag-of-Words].

Das daraus resultierende Vokabular ist eine Liste der eindeutigen Token, die die Grundlage für die Verarbeitung des Textes bildet. Es spielt eine zentrale Rolle, da es bestimmt, wie gut ein Modell den Text analysieren und verstehen kann. Die Grösse des Vokabulars kann stark variieren und hängt von der Art des Textes und der gewählten Tokenisierungsmethode ab. Ein kleines Vokabular kann bei einfachen Aufgaben oder spezialisierten Texten effizient sein, während für komplexe Sprachmodelle häufig ein sehr grosses Vokabular verwendet wird, das Millionen von Token umfasst. Vokabulare bieten eine erste Möglichkeit, Texte zu analysieren. Allerdings wird dabei lediglich eine Liste von eindeutigen Wörtern bereitgestellt. Eine inhaltliche Gewichtung ist dabei nicht möglich [16, Kap. Representing Language as a Bag-of-Words].

4.1.3 Bag-of-Words

Das Modell *Bag-of-Words* (BoW) wird verwendet, um anschliessend die Häufigkeit der Token zu quantifizieren und in einem Vektor darzustellen. Im untenstehenden Beispiel wird der Satz *My cat is cute* analysiert. Für jedes Wort im zuvor erstellten Vokabular wird gezählt, wie oft es im Text vorkommt. Dies führt zu einem Vektor, der in diesem Fall die Werte [0, 1, 0, 1, 0, 1, 1] enthält. Jeder Wert steht dabei für die Häufigkeit eines bestimmten Worts im Text. Diese numerische Darstellung ermöglicht es, Texte mathematisch zu analysieren und mit maschinellen Lernverfahren zu verarbeiten [16, Kap. Representing Language as a Bag-of-Words].

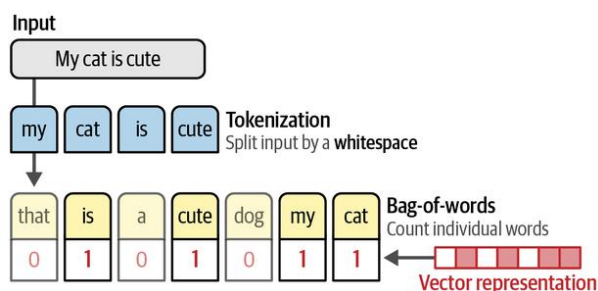


Abbildung 8: Funktionsweise des *Bag-of-Word-Modells* [16, Kap. Representing Language as a Bag-of-Words].

Das Modell hat einige Vorteile. Es ist leicht umsetzbar und dient als grundlegende Methode, um Texte in eine maschinenlesbare Form zu überführen. Dabei berücksichtigt es nicht nur, ob ein Wort existiert, sondern auch, wie oft es vorkommt, wodurch sich Worthäufigkeiten vergleichen lassen und besser festgestellt werden kann, welche Schwerpunkte im Text liegen. Diese Einfachheit bringt jedoch auch Einschränkungen mit sich. So wird die Reihenfolge der Wörter ignoriert, was den Verlust von Kontextinformationen zur Folge hat. Zudem entstehen bei grossen Vokabularen oft lange Vektoren, da viele Wörter in einem bestimmten Text nur wenig vorkommen. Semantische Zusammenhänge zwischen ähnlichen Wörtern wie *Auto* und *Fahrzeug* können ebenfalls nicht abgebildet werden, da diese Methode keine Bedeutungsbeziehungen berücksichtigt. Trotz dieser Limitationen ist es ein erster Ansatz, Textdaten für maschinelle Lernverfahren nutzbar zu machen und bildete die Basis für fortschrittlichere Modelle wie neuronale Netzwerke und Transformerrchitekturen [16, Kap. Representing Language as a Bag-of-Words].

4.1.4 Neuronale Netzwerke

Neuronale Netzwerke bilden die Grundlage vieler Deep-Learning-Methoden, die heute in der Textverarbeitung Anwendung finden. Die untenstehende Abbildung zeigt die Struktur eines neuronalen Netzwerks, das aus mehreren Schichten von Knoten besteht. Diese sind über Verbindungen miteinander verknüpft, die jeweils durch eine lineare Gleichung beschrieben werden. Die Verbindungen, auch als Gewichte bezeichnet, bestimmen die Stärke des Einflusses eines Knotens auf einen anderen. Durch das Training des Netzwerks werden diese Gewichte so angepasst, dass das Modell die bestmöglichen Ergebnisse liefert. Die Stärke neuronaler Netzwerke liegt in ihrer Fähigkeit, komplexe Muster und Beziehungen in den Daten zu erkennen – eine Schlüsselkomponente bei der Verarbeitung natürlicher Sprache. Ein weiterer Vorteil ist ihre Anpassungsfähigkeit, da sie mit unterschiedlichen Arten von Daten und Aufgaben umgehen können. Zudem ermöglichen sie es, verborgene Zusammenhänge in grossen Datenmengen zu identifizieren, die für den Menschen oft nicht direkt erkennbar sind. Die nachfolgende Abbildung zeigt einen Mechanismus zur Erkennung unerwünschter E-Mails (Spam). Inhalte (Features) werden gewichtet, um daraus eine Wahrscheinlichkeit zu berechnen, mit der die E-Mail als Spam klassifiziert wird. Hier sei darauf hingewiesen, dass auch fehlerhafte Ergebnisse (false positives) auftreten können, bei denen legitime E-Mails fälschlicherweise als Spam markiert werden. Dies kann entsprechend zu einer Verzerrung führen [16, Kap. Better Representations with Dense Vector Embeddings].

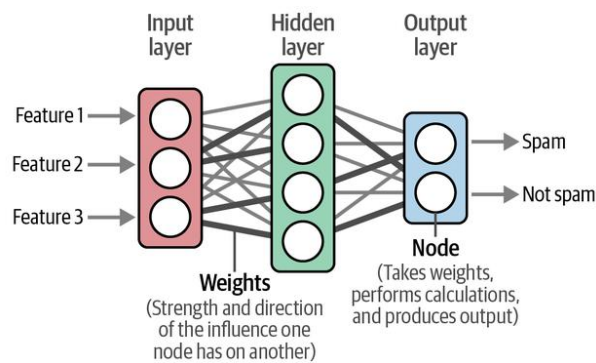


Abbildung 9: Funktionsweise eines neuronalen Netzwerks [16, Kap. Better Representations with Dense Vector Embeddings].

Ein wesentlicher Fortschritt in der Textverarbeitung war die Einführung von sogenannten Embeddings. Es handelt sich dabei um mathematische Repräsentationen von Wörtern, Sätzen oder anderen Daten, die sie in einem mehrdimensionalen Raum als Punkte abbilden lassen, um deren Bedeutung und Beziehungen besser zu verstehen. Dieser Ansatz löst viele Probleme älterer Methoden, wie beispielsweise dem *Bag-of-Words-Modell*. Im multidimensionalen Raum wird jedes Wort als Vektor repräsentiert, wobei jede Dimension eine bestimmte Eigenschaft oder ein Merkmal eines Wortes darstellt. Diese Werte geben an, wie stark ein Wort mit einem bestimmten Merkmal assoziiert ist. Dabei bewertet das Modell den Zusammenhang eines Wortes mit einer Vielzahl von Merkmalen, welche durch den Trainingsprozess des Modells bestimmt werden. Ein Vektor für das Wort *Katze* könnte beispielsweise stark mit Eigenschaften wie *Tier* oder *klein* korrelieren, während das Wort *Apfel* stärker mit *Frucht* oder *Essen* in Verbindung gebracht wird. Die beiden nachfolgenden Abbildungen sind stark vereinfacht, sollen aber die Funktionsweise veranschaulichen. Die erste illustriert, wie ein Modell die Wahrscheinlichkeit berechnet, dass ein Wort einer bestimmten Eigenschaft zugeordnet wird. Für das Wort *cat* wird beispielsweise eine sehr hohe Wahrscheinlichkeit (91%) für das Merkmal *animal* berechnet. Das bedeutet, dass das Modell *cat* stark mit dieser Eigenschaft verknüpft. Weitere Eigenschaften wie *plural* werden ebenfalls berücksichtigt, was zeigt, dass das Modell grammatikalische und kontextuelle Informationen lernen kann [16, Kap. Better Representations with Dense Vector Embeddings].

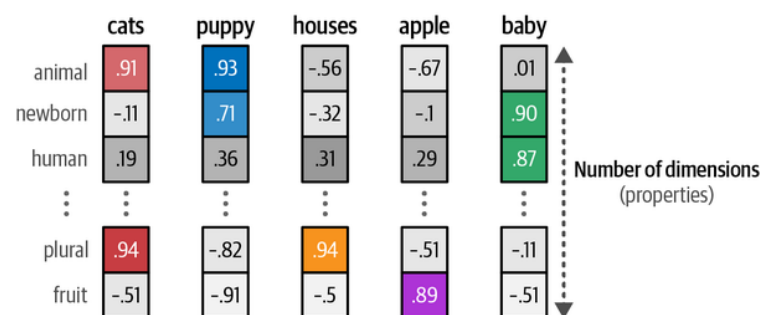


Abbildung 10: Wahrscheinlichkeiten für Wortmerkmale [16, Kap. Better Representations with Dense Vector Embeddings].

Der Vorteil von Embeddings liegt in ihrer Fähigkeit, semantische Ähnlichkeiten zu erkennen. Wörter, die ähnliche Bedeutungen haben oder in ähnlichen Kontexten verwendet werden, sind im multidimensionalen Raum näher beieinander. Die nächste Abbildung zeigt, dass die Wörter *cat* und *dog* ähnliche Vektoren haben, während *apple* und *fruit* ebenfalls nah beieinander liegen, jedoch in einer anderen semantischen Region. Daraus lassen sich n-dimensionale Darstellungen aufspannen, wobei mithilfe von Distanzmassen berechnet werden kann, wie die Wörter zueinander in Beziehung stehen [16, Kap. Better Representations with Dense Vector Embeddings]:

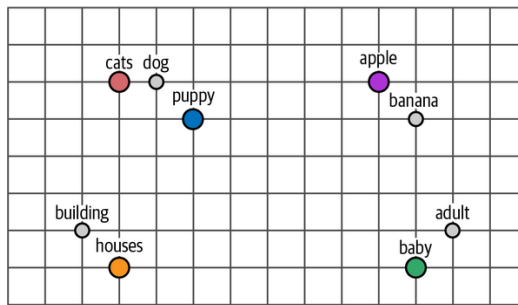


Abbildung 11: Visuelle Darstellung von Wortbeziehungen im zweidimensionalen Raum [16, Kap. Better Representations with Dense Vector Embeddings].

Bei der technischen Darstellung von Sprache werden nicht nur Wörter berücksichtigt, sondern auch abstraktere Ebenen wie Bedeutungsbeziehungen und Kontextinformationen. Diese Ebenen repräsentieren oft komplexe Muster, die für uns Menschen schwer nachvollziehbar sind. Die nachstehende Abbildung illustriert anhand eines einfachen Beispiels, wie unterschiedliche Daten zu mehreren Embeddings führen können. Diese repräsentieren verschiedene Facetten der semantischen und syntaktischen Bedeutung, die das Modell lernt und nutzt [16, Kap. Types of Embeddings].

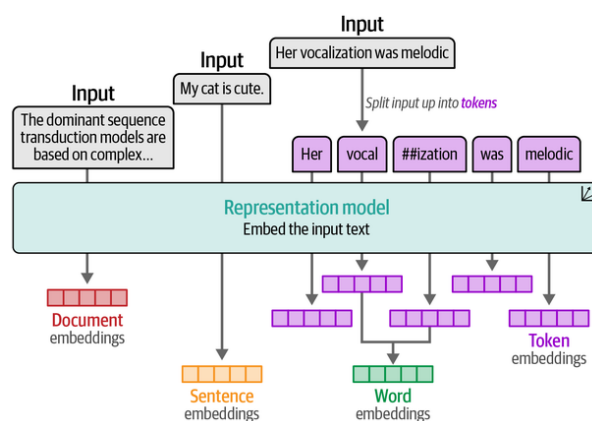


Abbildung 12: Daten führen zu unterschiedlichen Beziehungen im Modell [16, Kap. Types of Embeddings].

Neuronale Netzwerke bieten somit eine flexible Grundlage für die Verarbeitung natürlicher Sprache. Allerdings stoßen sie bei der Verarbeitung von Sequenzen, wie sie in Texten häufig vorkommen, an ihre Grenzen. Insbesondere fehlt es ihnen an der Fähigkeit, zeitliche Abhängigkeiten oder den Kontext zwischen aufeinanderfolgenden Wörtern über längere Sequenzen hinweg zu berücksichtigen. Um diese Herausforderung zu lösen, wurden rekurrente neuronale Netzwerke (RNNs) entwickelt. Diese Modelle sind darauf spezialisiert, Informationen aus früheren Sequenzen in ihren Berechnungen zu berücksichtigen. Ein Beispiel ist die Analyse des Begriffs *System*, der je nach Kontext unterschiedliche Bedeutungen hat, etwa als technisches (Computer-)System oder als Familiensystem in der Sozialen Arbeit. RNNs ermöglichen es, Zusammenhänge in solchen Sequenzen zu verstehen, indem sie vorherige Informationen speichern und in den Entscheidungsprozess einfließen lassen [16, Kap. Encoding and Decoding Context with Attention].

4.1.5 Rekurrente neuronale Netzwerke

Rekurrente neuronale Netzwerke (RNNs) stellen einen wichtigen Fortschritt in der Verarbeitung von Sprachsequenzen dar. Sie können die Reihenfolge von Wörtern in Texten berücksichtigen, was zu einem besseren Textverständnis führt. Wie in der untenstehenden Abbildung dargestellt, kommen RNNs beispielsweise bei Übersetzungsaufgaben zum Einsatz. Am Beispiel wird erklärt, wie der Satz wie *I love llamas* ins Niederländische übersetzt wird. Der Prozess besteht aus zwei Hauptschritten: der Kodierung (Encoding) und der Dekodierung (Decoding). Im Kodierungsschritt werden die

Eingabewörter nacheinander verarbeitet, wobei jede Wortsequenz in einen Kontextvektor (Embedding) umgewandelt wird, der die gesamte Eingabe repräsentiert. Dieses Kontext-Embedding dient als Grundlage für den Dekodierungsschritt, bei dem der Ausgabesatz generiert wird [16, Kap. Encoding and Decoding Context with Attention].

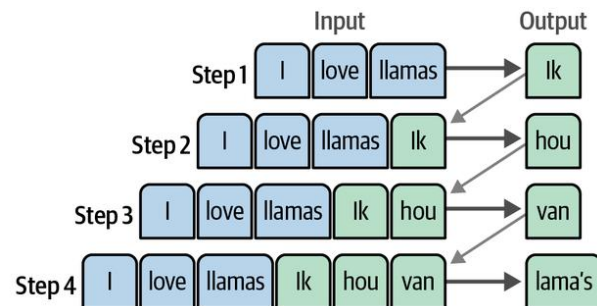


Abbildung 13: Autoregressive Abfragestruktur eines RNNs [16, Kap. Encoding and Decoding Context with Attention].

Ein wesentliches Merkmal von RNNs ist ihre autoregressive Natur: Jedes Wort wird basierend auf den zuvor generierten Wörtern erzeugt. Dies führt jedoch zu einer sequenziellen Verarbeitung, die die Parallelisierung während des Trainings verhindert und den Rechenaufwand erheblich steigert. Insbesondere bei längeren Sätzen kann diese Architektur ineffizient sein, da der gesamte Kontext in einem einzigen Vektor gespeichert wird, was den Umgang mit umfangreichen Texten erschwert [16, Kap. Encoding and Decoding Context with Attention]. Um diese Einschränkungen zu überwinden, wurden Mechanismen wie *Attention* entwickelt. Diese erlauben, gezielt auf wichtige Teile der Eingabe zu fokussieren, indem für jedes Wort im Eingabesatz relevante Verbindungen zu den Wörtern im Ausgabesatz hergestellt werden.

4.1.6 Transformatoren

Die Transformer-Architektur, die durch die Arbeit *Attention is all You Need* von Vaswani et al. bekannt wurde, markierte einen Wendepunkt in der Textverarbeitung. Diese Architektur basiert vollständig auf dem Konzept der Selbstaufmerksamkeit (Self-Attention), wodurch die Beziehungen zwischen Wörtern unabhängig von ihrer Position analysiert werden können [17]. Transformer-Modelle zeichnen sich durch ihre Fähigkeit aus, komplexe Texte vollständig zu analysieren und dabei auch den gesamten Kontext zu erfassen. Dies ist insbesondere für Aufgaben wie Textgenerierung und maschinelle Übersetzung von entscheidender Bedeutung. Ein weiterer Vorteil der Transformer-Architektur ist die parallele Verarbeitung von Eingaben, was die Trainingszeit deutlich reduziert und die Skalierung auf sehr grosse Datensätze ermöglicht. Transformer-Modelle bilden die Basis für sogenannte Large Language Models (LLMs), die komplexe Textverarbeitungsaufgaben mit hoher Effizienz und Genauigkeit bewältigen können. Diese Modelle wurden auf umfangreichen Textkorpora trainiert und nutzen Konzepte wie Attention-Mechanismen und Embedding-Modelle, um kontextuelle und semantische Zusammenhänge in Texten zu erfassen. Dabei unterscheiden sich die Modelle in drei Architekturvarianten: Encoder-Modelle, Decoder-Modelle und hybride Encoder-Decoder-Modelle. Eine kurze Beschreibung dieser Architekturen und deren entsprechende Einsatzgebiete werden in der nachfolgenden Tabelle zusammengefasst [18].

Typ	Bekannte Modelle	Beschreibung	Einsatzgebiete
Encoder	BERT	Die Encoder-Architektur wird für das tiefgehende Verstehen von Texten verwendet. Sie erzeugt eine kontextuelle Vektordarstellung der Eingabesequenz, die eine tiefere Analyse der Wortbeziehungen und Bedeutungen innerhalb des Textes ermöglicht. Dies ist besonders nützlich für Aufgaben wie Stimmungsanalyse, thematische Einordnung oder die Erkennung und Klassifizierung von Eigennamen und Schlüsselbegriffen mittels Named Entity Recognition (NER).	Textklassifikation, Stimmungsanalyse, NER

Decoder		<i>BERT</i> verwendet ausschliesslich Selbstaufmerksamkeit, um alle relevanten Teile der Eingabe gleichzeitig zu berücksichtigen, ohne eine neue Sequenz zu generieren. Dies macht es ideal für Aufgaben, die eine präzise Analyse des vorhandenen Textes erfordern. Die Modelle können Texte zusammenfassen, allerdings würde das Modell eine extraktive anstatt generative Zusammenfassung erstellen.	
	GPT-3	Der Decoder generiert neuen Text basierend auf einer Eingabeaufforderung oder einem Kontext. Diese Architektur bildet die Grundlage für GenAI, auch bekannt als generative AI. Diese autoregressive Eigenschaft bedeutet, dass jedes neu generierte Wort nur auf den vorherigen Wörtern basiert. Diese Architektur ist ideal für Sprachgenerierung, wie etwa das Verfassen von Geschichten, das Beantworten von Fragen in natürlicher Sprache oder die Verwendung in Chatbots. Das Modell <i>GPT-3</i>, das 2020 von <i>OpenAI</i> eingeführt wurde, ist ein leistungsfähiges Sprachmodell, das menschenähnliche Antworten in einer Vielzahl von Themengebieten generieren kann. Die Modelle können Texte zusammenfassen, haben ihren Fokus aber mehr auf dem Generieren von Text, anstatt den ganzen Eingabetext zu erfassen, was potenziell zu weniger präzisen Antworten führen kann.	Textgenerierung, Chatbots, Fragebeantwortung
Encoder-Decoder	T5, M2M-100	Der Encoder nimmt eine Eingabesequenz auf und erzeugt eine Vektordarstellung, die an den Decoder übergeben wird. Der Decoder nutzt sowohl Selbstaufmerksamkeit, um wichtige Teile der bereits erzeugten Ausgabesequenz zu erkennen, als auch gegenseitige Aufmerksamkeit, um die relevanten Informationen aus der Encoder-Ausgabe zu verwenden. Diese Architektur wird häufig für Aufgaben wie Sprachübersetzungen verwendet, bei denen das Modell den Kontext eines Satzes in einer Sprache analysiert und in eine andere Sprache übersetzt. Beispiele sind <i>Google Translate</i>, welches das <i>T5-Modell</i> verwendet, und <i>Facebooks M2M-100</i>, das Übersetzungen zwischen 100 verschiedenen Sprachen ermöglicht. Für die Zusammenfassung von Text sind diese Modelle optimal geeignet.	Sprachübersetzung, Text-zu-Text-Transformation

Tabelle 1: Architektur-Varianten von Transformatoren [18].

Die Leistungsfähigkeit von Transformer-Modellen hängt nicht nur von ihrer Architektur ab, sondern massgeblich von der Qualität und Menge der Trainingsdaten, die zur Erzeugung hochdimensionaler Embeddings und zur Berechnung von Wahrscheinlichkeitsverteilungen genutzt werden. Ebenso beeinflusst das Trainingsverfahren direkt die Fähigkeit eines Modells, kontextuelle und semantische Zusammenhänge zu erfassen. Das folgende Kapitel beleuchtet daher die Trainingsprozesse und zeigt, wie grosse Sprachmodelle durch gezielte Feinabstimmung (Fine Tuning) für spezifische Anwendungsfälle optimiert werden.

4.1.7 Trainieren von Modellen

Damit Transformer-Modelle ihr volles Potenzial entfalten können, müssen sie auf umfangreichen Datensätzen trainiert werden. Die Qualität und Menge der Trainingsdaten sowie die Skalierung der Modellparameter spielen dabei eine entscheidende Rolle. Ein frühes Beispiel für ein gross angelegtes Trainingsverfahren ist das Modell *GPT-1*, das als Grundlage für spätere leistungsfähigere Modelle diente. Das Modell wurde auf einem Datenkorpus bestehend aus 7.000 Büchern und *Common Crawl* trainiert. Letzteres ist eine gemeinnützige Organisation, die das Internet durchsucht und ihre Archive sowie Datensätze der Öffentlichkeit kostenlos zur Verfügung stellt [19]. Das daraus entstandene

Modell verfügte über 117 Millionen Parameter. Auf Basis der heutigen technischen Fortschritte wird angenommen, dass grössere Modelle leistungsfähiger sind. Die folgende Grafik verdeutlicht die Entwicklung der Parameteranzahl von GPT-Modellen in den letzten Jahren und deren Versionen [16, Kap. Generative Models: Decoder-Only Models]:

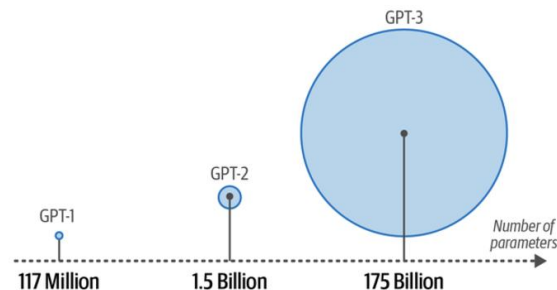


Abbildung 14: Anzahl Parameter bei GPT-Versionen [16, Kap. Generative Models: Decoder-Only Models].

Ein entscheidender Vorteil der modernen Modelle liegt nicht nur in ihrem zunehmenden Verständnis von Sprache, sondern auch in der Möglichkeit, sie durch gezielte Feinabstimmung für spezifische Aufgaben zu optimieren. Die beiden nachfolgenden Abbildungen veranschaulichen die Grundprinzipien dieses Prozesses. Die erste Abbildung zeigt ein allgemeines Schema des Trainings: Zunächst wird ein Modell mithilfe eines umfangreichen Datensatzes vortrainiert. Anschliessend werden spezifische, oftmals überwachte Daten verwendet, um das Modell gezielt zu verfeinern. Dieser Ansatz führt zu einem spezialisierten und optimierten Modell, das für konkrete Anwendungsfälle eingesetzt werden kann [16, Kap. Language Models].

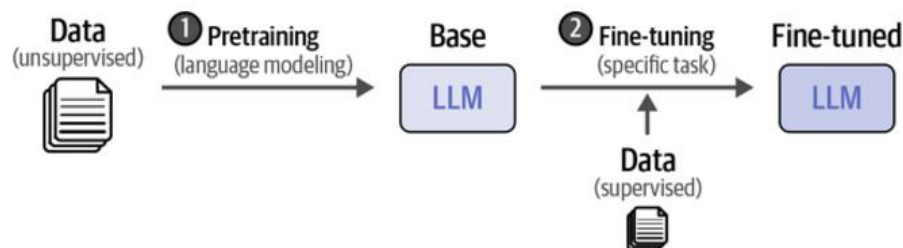


Abbildung 15: Grundschemata für das Modelltraining [16, Kap. Language Models].

Die nächste Abbildung veranschaulicht das Trainings- und Optimierungsverfahren am Beispiel des Modells BERT. Zunächst wird mit einem umfangreichen Trainingskorpus vortrainiert, der beispielsweise aus Wikipedia-Daten besteht. Dieser Schritt ermöglicht es dem Modell, grundlegende Sprachmuster, grammatikalische Strukturen und semantische Zusammenhänge zu erlernen. Anschliessend wird das vortrainierte Modell durch Feinabstimmung an spezifische Anwendungsfälle angepasst, wie z. B. Textklassifikation, Named Entity Recognition oder Paraphrasenerkennung. Dabei kommen domänenspezifische oder überwachte Daten zum Einsatz, um die Leistung für zielgerichtete Aufgaben zu maximieren [16, Kap. Representation Models: Encoder-Only Models]. Diese zweistufige Herangehensweise kombiniert die Vorteile eines generalisierten Sprachverständnisses mit einer hohen Spezialisierung für konkrete Anwendungen, wie beispielsweise spezifisches Domänenwissen in der Sozialen Arbeit.

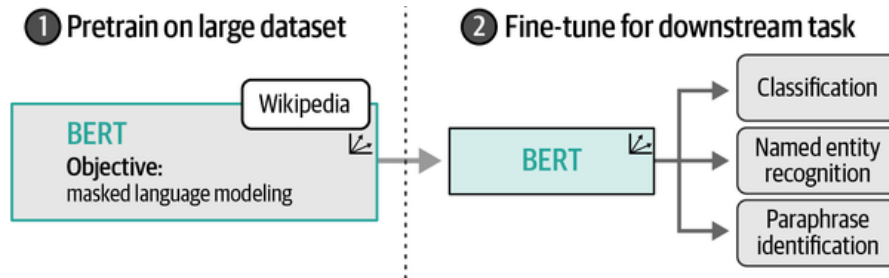


Abbildung 16: Optimierung des Modells BERT für spezifische Anwendungsfälle [16, Kap. Representation Models: Encoder-Only Models].

Ableitend für die Masterarbeit wird die Auswahl des richtigen Modells entscheidend sein. Neben der allgemeinen Optimierung durch Feinabstimmung muss speziell untersucht werden, welche Modellvariante sich am besten für die Aufgabe der Berichterstellung eignet. Eine zentrale Herausforderung dabei besteht im Umgang mit sensiblen Daten. Da die Trainingsdaten häufig persönliche und schützenswerte Informationen enthalten, müssen Massnahmen ergriffen werden, um zu verhindern, dass diese in das Modell einfließen.

Halevy et al. beschreiben eine weitere zentrale Herausforderung beim Trainieren von Modellen: Für viele Aufgaben existieren keine grossen, frei verfügbaren Korpora, die für ein spezifisches Modelltraining genutzt werden können. In solchen Fällen ist eine qualifizierte menschliche Annotation unerlässlich, da ohne sie keine verlässliche Grundlage für das Modelltraining besteht [20, S. 2]. Der Begriff Annotation bezeichnet die manuelle oder automatisierte Markierung von Textdaten, um beispielsweise Begriffe, Entitäten oder Satzstrukturen eindeutig zu kennzeichnen. Dies gilt insbesondere für Modelle, die Domänenwissen erfordern – beispielsweise in der Sozialen Arbeit – und daher auf spezifisch annotierte Daten angewiesen sind. Um diese Problematik zu adressieren, kann die Nutzung eines sogenannten *Golden Data Sets* eine Lösung bieten. Dieses besteht aus einer kleinen, vollständig annotierten Datenmenge, die als Referenz (Ground Truth) dient.

Wie Alamar und Grootendorst zeigen, wird ein solches Set beispielsweise im Augmented SBERT-Verfahren genutzt, um ein hochpräzises Modell zu trainieren. Dieses Modell kann anschliessend grössere, unannotierte Datensätze (*Silver Data Sets*) automatisch labeln, wodurch das verfügbare Trainingsmaterial erheblich erweitert wird. Die nachfolgende Grafik verdeutlicht dieses Konzept und zeigt strategische Ansätze auf, wie mit begrenzten Daten ein leistungsfähiges Modell entwickelt werden kann. Obwohl in dieser Arbeit nicht auf die methodische Umsetzung eines *Golden Data Sets* eingegangen wird, sind solche Verfahren ein relevanter Forschungsansatz zur Optimierung von Modellen mit begrenzten Datenressourcen [16, Kap. Augmented SBERT].

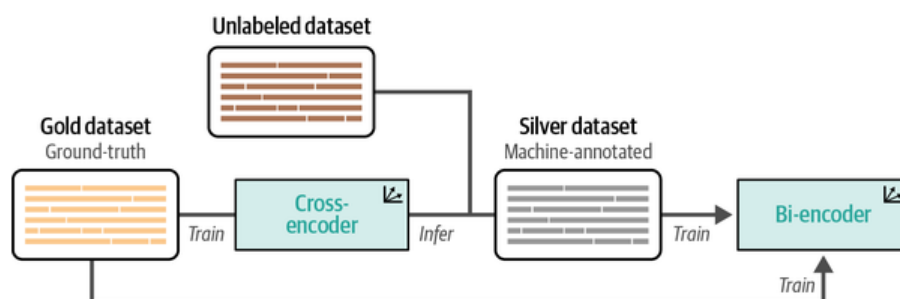


Abbildung 17: Nutzung eines Golden Data Sets, um nicht gelabelte Daten automatisch zu annotieren [16, Kap. Augmented SBERT].

Während das Training und die Optimierung von Modellen durch Verfahren wie Feinabstimmung und *Golden Data Sets* entscheidend für ihre Leistungsfähigkeit sind, bleibt eine zentrale Herausforderung bestehen: der Umgang mit sensiblen Daten. Eine vielversprechende Alternative zur direkten Modellintegration dieser Informationen ist Retrieval Augmented Generation (RAG). Dieses Verfahren ermöglicht es, externe Informationen und Datenquellen gezielt in den Prompt-Prozess einzubinden, sodass relevante Inhalte zur Laufzeit abgerufen werden, ohne dauerhaft in das Modell einzugehen. Das folgende Kapitel gibt einen kurzen Überblick über die Funktionsweise von RAG und zeigt, welche Vorteile es für den sicheren Einsatz von KI bei sensiblen Daten bietet [16, Kap. From Search to RAG].

4.1.8 Retrieval Augmented Generation

Ein weiterer Ansatz besteht darin, dem Modell während der Laufzeit zusätzliche oder sensitive Informationen bereitzustellen. Dies schützt sicherheitskritische und schützenswerte Daten wie sensible Journaldaten vor einer direkten Modellintegration. Retrieval Augmented Generation (RAG) folgt einem mehrstufigen Prozess, der in der nachfolgenden Abbildung veranschaulicht wird [16, Kap. From Search to RAG].



Abbildung 18: Funktionsweise von RAG [16, Kap. From Search to RAG].

1. **Eingabeaufforderung (Prompt)**
Die Nutzereingabe wird zunächst in eine strukturierte Anfrage umgewandelt, die das Modell versteht.
2. **Embedding-Modell**
Die Anfrage wird von einem Embedding-Modell in eine Vektorrepräsentation umgewandelt, die es ermöglicht, semantische Ähnlichkeiten zwischen der Anfrage und bereits vorhandenen Informationen zu bewerten.
3. **Durchsuchen der Datenbank oder Quelle**
Mithilfe der Vektorrepräsentation werden relevante Kontexte aus einer externen Datenbank oder Quelle identifiziert, die von der Institution bereitgestellt wird. Diese kann strukturierte oder unstrukturierte Informationen enthalten, wie beispielsweise Fachkonzepte, ethische Grundlagen oder Journaldaten. Sensible Informationen werden dabei durch den RAG-Prozess gefiltert und gezielt abgerufen, ohne dauerhaft in das Modell integriert zu werden.
4. **Relevanter Kontext**
Die durch den Vergleich der Einbettungen identifizierten Informationen werden ausgewählt und der ursprünglichen Eingabeaufforderung hinzugefügt.
5. **Aktualisierte Eingabeaufforderung**
Die angereicherte Eingabeaufforderung wird schliesslich an das Sprachmodell weitergegeben, welches daraufhin eine optimierte und kontextsensitive Antwort generiert.

Durch die Kombination von vortrainierten Modellen mit spezialisierten, institutionell bereitgestellten Datenquellen bietet RAG zahlreiche Vorteile:

- Es wird die Privatsphäre gewahrt, da sensible Informationen wie Journaldaten nicht Teil des Modells werden, sondern sicher in externen Quellen bleiben.
- Der Ansatz ermöglicht eine flexible Anpassung an verschiedene Aufgaben, ohne dass die zusätzlichen Informationen im Vorfeld trainiert werden müssen.
- RAG ist ressourcenschonend und effektiv, da vortrainierte Modelle in Kombination mit externen Datenbanken präzise und kontextsensitive Ergebnisse liefern können.
- Institutionelle Daten, wie Journaldaten oder fachspezifische Konzepte, können gezielt und sicher gefiltert in die Laufzeitanfragen einfließen.

Mit diesem Abschnitt endet die Betrachtung der grundlegenden Prinzipien der KI für die Textverarbeitung. Aufbauend auf diesen Grundlagen richtet sich der Fokus nun auf die praktische Anwendung von KI-Modellen. Das folgende Kapitel analysiert zentrale Aspekte der Modellnutzung und beleuchtet, welche methodischen, datenbezogenen und technischen Herausforderungen berücksichtigt werden müssen, um KI-gestützte Systeme effizient und zielgerichtet einzusetzen.

4.2 Modelle in der praktischen Anwendung

Der erfolgreiche Einsatz von KI in der Praxis erfordert eine präzise Abstimmung verschiedener Faktoren. Neben der Wahl geeigneter Architekturen spielen die Qualität der Trainingsdaten, die Gestaltung von Prompts und die Evaluierungsmethoden eine entscheidende Rolle für die Genauigkeit und Zuverlässigkeit der generierten Ergebnisse. Zudem gewinnen autonome Agentensysteme (Agent AI) an Bedeutung, da sie komplexe Aufgaben eigenständig bearbeiten und bestehende Modelle gezielt erweitern können. Gleichzeitig wird die Erklärbarkeit von KI-Systemen (Explainable AI) immer wichtiger, um Entscheidungen nachvollziehbar zu machen und Vertrauen in die Technologie zu stärken. Dieses Kapitel beleuchtet die zentralen Aspekte der praktischen Anwendung von KI-Systemen.

4.2.1 Modellauswahl

Alammar und Grootendorst schätzen, dass im September 2024 etwa 800'000 Modelle mit diversen Anwendungen, von grossen Sprachmodellen bis hin zu Modellen für Audio- und Tabellendaten, verfügbar sind. Sie betonen, dass die Plattform Hugging Face [21] nahezu alle Open-Source-LLMs zum Download anbietet [16, Kap. Generating Your First Text]. Zusätzlich schätzen sie, dass etwa 60'000 Modelle speziell für Textklassifizierungsaufgaben und über 8'000 Modelle zur Generierung von Embeddings entwickelt wurden [16, Kap. Model Selection]. Die untenstehende Tabelle vergleicht Modelle wie BERT, GPT-3 oder T5 anhand zentraler Merkmale, darunter der Pre-Training-Ansatz, die Anzahl der Parameter, der Zugangsmöglichkeit (API oder Open-Source) und der Eignung für verschiedene Aufgaben wie Textklassifikation, Konversations-KI oder Maschinelle Übersetzung. So ist beispielsweise GPT-3 mit 175 Milliarden Parametern auf Content-Generierung spezialisiert, während BERT aufgrund seines Ansatzes besonders für Textklassifikationen geeignet ist. Die Farbskala der nachfolgenden Abbildung verdeutlicht, wie gut ein Modell für die jeweilige Aufgabe geeignet ist [22].

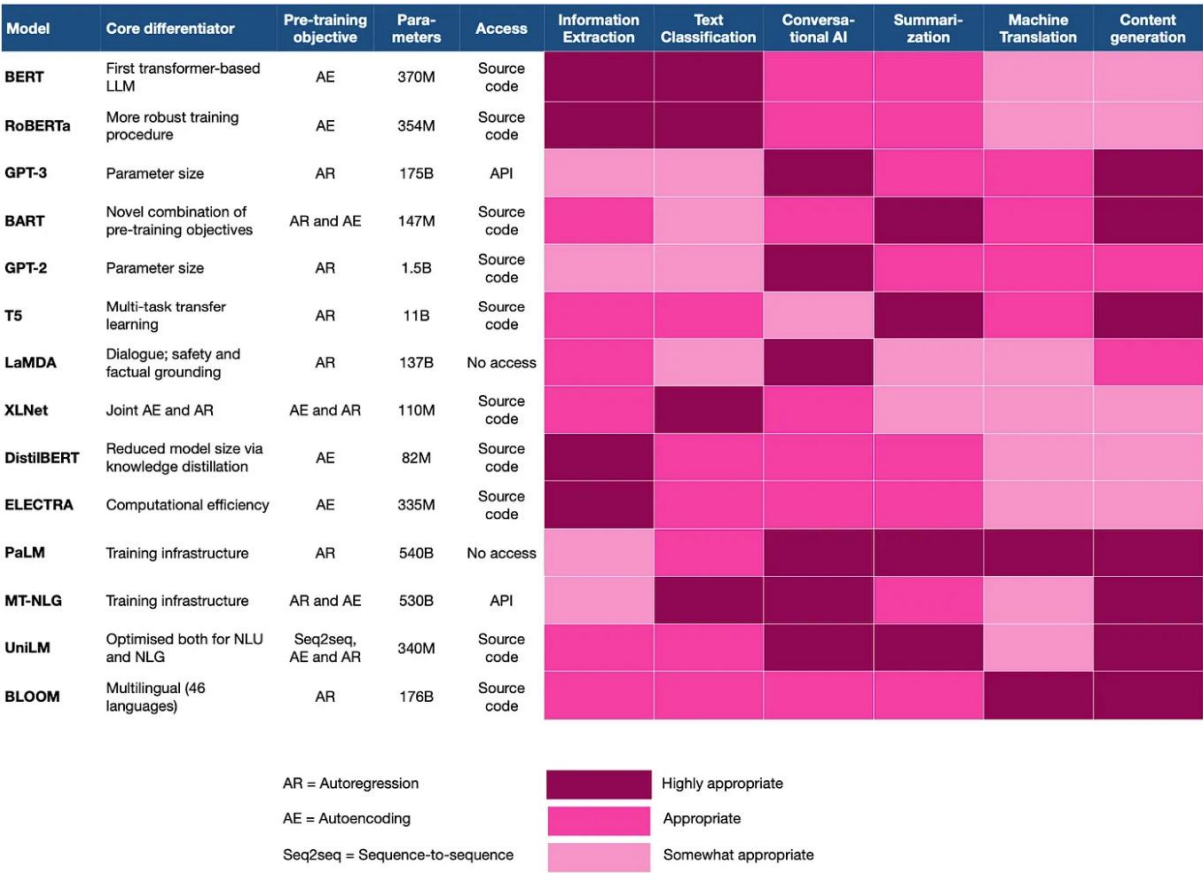


Abbildung 19: Zusammenfassung der Merkmale beliebter Sprachmodelle [20].

Neben der Eignung für spezifische Aufgaben stehen zahlreiche Benchmark-Tests zur Verfügung, die die Leistungsfähigkeit von Modellen messen und vergleichen. Die nachfolgende Grafik zeigt die Ergebnisse des Massive Multitask Language Understanding (MMLU), eines Benchmarks, der bewertet, wie gut Modelle Sprache verstehen. Dabei werden verschiedene Aspekte berücksichtigt, darunter das Kontextverständnis, die Verarbeitung von Token und der Umfang des abgedeckten Wissens [23].

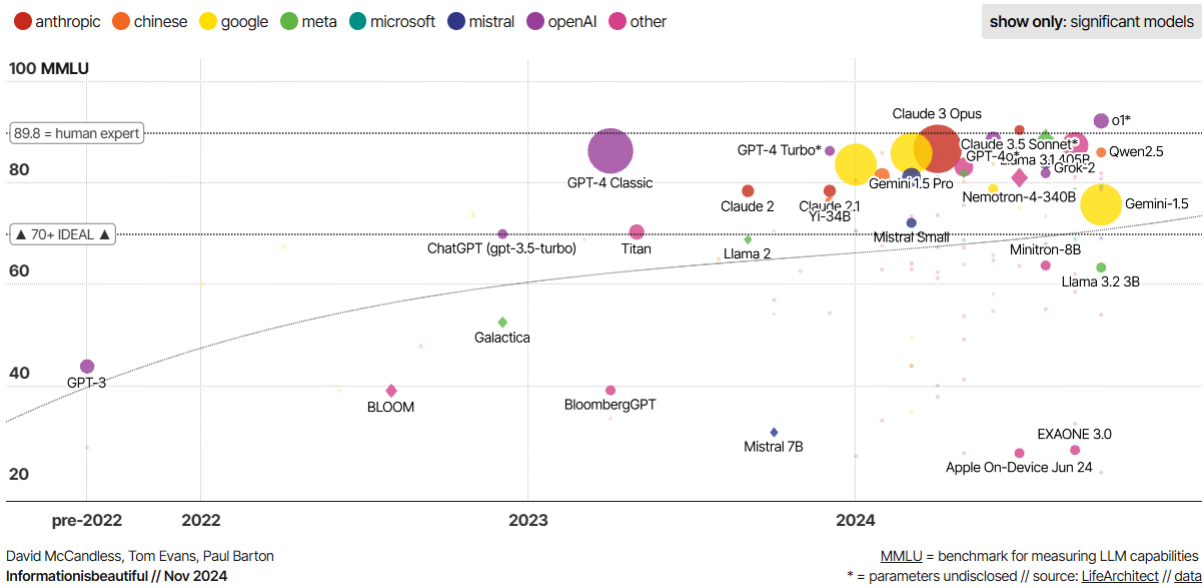


Abbildung 20: Entwicklung der MMLU-Leistung bedeutender Sprachmodelle (2022-2024) [23].

4.2.2 Trainingsdaten und ihre Implikationen

Ein Modell ist immer nur so gut wie die Daten, mit denen es trainiert wurde – daher spielt die Qualität der Trainingsdaten eine entscheidende Rolle. Nach Strasser und Niedermayer beeinflusst sie massgeblich die Präzision, Fairness und kulturelle Eignung der Ergebnisse. Unausgewogene oder einseitige Daten können bestehende Vorurteile verstärken, während veraltete Informationen die Aktualität und Relevanz der generierten Inhalte erheblich mindern. Eine vielseitige und aktuelle Datengrundlage ist daher essenziell, um präzise und faire Ergebnisse sicherzustellen [3, S. 121 ff.].

Nachstehende Grafik zeigt eine Interaktion mit einem älteren *GPT-3.5-Turbo-0125-Modell*, das nach dem aktuellen Monarchen Grossbritanniens gefragt wird. Das Modell ist nicht in der Lage, zu erkennen, dass Queen Elizabeth II am 8. September 2022 gestorben und König Charles III ihr Nachfolger ist, da diese Informationen nicht im Trainingsdatensatz enthalten sind [24].

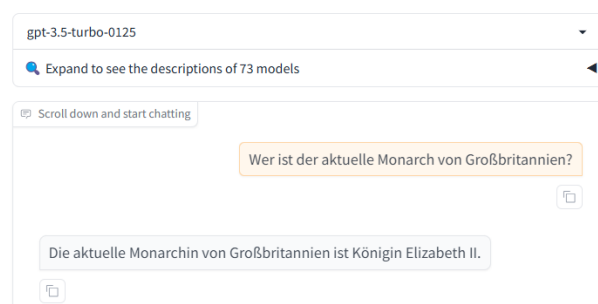


Abbildung 21: Veraltete Trainingsdaten können zu falschen Antworten führen [24].

Das Beispiel zeigt, dass die Aktualität der Trainingsdaten ein entscheidender Faktor für die Genauigkeit der Antworten eines Modells darstellt. Obwohl moderne Implementationen wie beispielsweise bei *GPT-4o* [25] in der Lage sind, Online-Recherchen durchzuführen und diese Inhalte für Antworten zu verwenden, bleibt die regelmässige Aktualisierung der Trainingsdaten wichtig, um die Relevanz und Präzision zu gewährleisten. Neben der Aktualität sind jedoch auch noch weitere Faktoren entscheidend, um Verzerrungen und Fehler zu vermeiden:

Kulturelle Vielfalt

- Die UNESCO hebt hervor, wie wichtig es ist, kulturelle Vielfalt in KI-Systemen zu fördern, um allen, einschliesslich lokaler Gemeinschaften, einen gleichberechtigten Zugang zu KI-Anwendungen zu ermöglichen [4, S. 20]. Diese Aspekte sind entscheidend, um die Nutzung von KI global gerecht zu gestalten. Gleichzeitig betonen Zwingmann und Gärtner, dass historische Daten oft unvollständig, verzerrt oder fehlerhaft sind, was die Ergebnisse der Modelle direkt beeinflussen kann [3, S. 240]. Auch Alammari und Grootendorst betonen die Bedeutung mehrsprachiger Modelle in der Praxis. Sie weisen darauf hin, dass Modelle, die überwiegend auf englischen Trainingsdaten basieren, zwar in englischsprachigen Kontexten besonders leistungsfähig sind, jedoch für Anwendungen in anderen Sprachen weniger geeignet sein können [16, Kap. Token Embeddings].

Transparenz und Fairness

- Ein weiterer relevanter Unterschied besteht zwischen Open-Source-Modellen und Closed-Source-Modellen. Open-Source-Modelle wie *BLOOM* bieten umfassende Einsicht in die Modellarchitektur und insbesondere auch in die verwendeten Trainingsdaten, was die Nachvollziehbarkeit und Transparenz deutlich verbessert. Im Gegensatz dazu behalten Closed-Source-Modelle wie *GPT-4* ihre Gewichte und Architektur unter Verschluss. Dies erschwert die Transparenz und beeinträchtigt das Vertrauen in die Ergebnisse solcher Modelle erheblich [16, Kap. Proprietary, Private Models]. Rigotti et al. weisen darauf hin, dass es bei KI-Anwendungen vermehrt Bedenken hinsichtlich Privatsphäre und Fairness gibt. Besonders problematisch sei der Mangel an Transparenz, der dazu führt, dass KI soziale Ausgrenzung und Diskriminierung begünstigen kann. Häufig reproduzieren Modelle Vorurteile, die bereits in ihren

Trainingskorpora enthalten sind. Ein bekanntes Beispiel hierfür ist das inzwischen eingestellte Rekrutierungstool von Amazon, das nachweislich Frauen systematisch diskriminiert hat [26]. Auch Kreissl und von Laufenberg merken an, dass Modelle zuverlässig Verzerrungen reproduzieren, wenn in den Trainingsdaten Stereotypen und kulturelle Vorurteile enthalten sind [13, S. 240].

Verzerrungen und Halluzinationen

- IBM beschreibt, dass KI-Modelle in ihren Daten Muster erkennen, die zu unpassenden, surrealen oder grundlegend falschen Antworten führen können. Dieses Phänomen wird als Halluzination bezeichnet und stellt eine zentrale Herausforderung bei der Generierung von KI-Inhalten dar. Um Halluzinationen zu minimieren, betont IBM die Bedeutung hochwertiger und relevanter Trainingsdatensätze. Zudem zeigt sich, dass Modelle, die ein domänenspezifisches Training durchlaufen haben, präzisere Ergebnisse liefern [27]. Auch Dziri et al. untersuchen dieses Problem und weisen darauf hin, dass neben der Architektur eines Modells insbesondere die Qualität der Trainingsdaten entscheidend für das Auftreten von Halluzinationen ist [28]. Darüber hinaus hebt Muhammad Zayyanu in seiner Untersuchung zu *BERT*, *GPT* und *T5* hervor, dass diese Modelle trotz ihrer beeindruckenden Leistungen weiterhin mit Verzerrungen in den Trainingsdaten, ethischen Herausforderungen bei der Inhaltsgenerierung und dem Risiko, bestehende Stereotype in Übersetzungen zu verstärken, konfrontiert sind [29, Kap. 6].

Während die Qualität der Trainingsdaten massgeblich die Leistungsfähigkeit eines Modells bestimmt, spielt auch die Art der Interaktion mit dem Modell eine entscheidende Rolle. Ein präziser und gezielter Einsatz von Prompts kann dazu beitragen, bestimmte Herausforderungen wie Verzerrungen oder Halluzinationen zu minimieren und die Ausgabequalität zu verbessern. Das folgende Kapitel beleuchtet daher die Bedeutung von Prompts und zeigt Methoden auf, um die Interaktion mit Sprachmodellen gezielt zu steuern.

4.2.3 Prompting

In diesem Kapitel wird der sogenannte Prompt als zentrales Element für die Interaktion mit einem LLM behandelt. Er stellt eine gezielte Anweisung oder einen Befehl dar, die dem Modell übermittelt werden, um spezifische Aufgaben zu initiieren und zielgerichtet zu lösen. Im vorangegangenen Kapitel wurden verschiedene Herausforderungen bei der Nutzung solcher Modelle aufgezeigt. Einige dieser Probleme können durch den gezielten und methodischen Einsatz von Prompts zumindest reduziert oder effizienter adressiert werden. Die nachfolgende Grafik illustriert einen einfachen Prompt, der zusätzlich mit relevanten Daten angereichert wurde [16, Kap. The Potential Complexity of a Prompt].

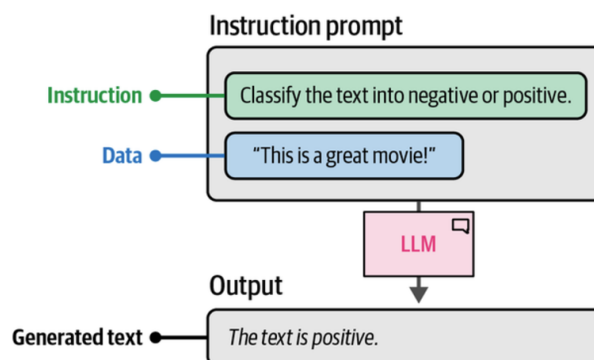


Abbildung 22: Das LLM wird mittels Prompt aufgefordert, einen Text zu klassifizieren [16, Kap. The Basic Ingredients of a Prompt].

Prompts können mehrere Anweisungen enthalten und verschiedene Techniken kombinieren. Eine effektive Gestaltung erfordert daher oft die gleichzeitige Anwendung mehrerer Methoden, welche in der nachfolgende Tabelle zusammengefasst werden. Diese stammen aus zwei Quellen und wurden durch

Beispiele aus der Praxis ergänzt [30, Kap. Overview of the Five Principles of Prompting ff.], [16, Kap. The Potential Complexity of a Prompt]:

Methode	Beschreibung	Beispiel
Richtung und Persona	Präzise Anweisungen, die das gewünschte Ziel und die Rolle des Modells klar definieren.	«Du bist ein Sozialpädagoge, der einen Bericht für Jugendliche mit psychischen Belastungen schreibt.»
Format angeben	Festlegung eines spezifischen Ausgabeformats, um konsistente und verwertbare Ergebnisse zu gewährleisten.	«Erstelle eine Checkliste mit fünf Punkten für die Umsetzung von barrierefreien Freizeitangeboten.»
Beispiele bereitstellen	Beispiele für gewünschte Ergebnisse, um Missverständnisse zu vermeiden und klare Erwartungen zu setzen.	«Ein Beispiel für ein erfolgreiches Programm ist ein inklusiver Musikworkshop mit professioneller Leitung.»
Qualität bewerten	Mechanismen zur Bewertung der generierten Inhalte, um die Effektivität von Prompts und Antworten zu verbessern.	«Bewerte die vorgeschlagenen Massnahmen nach Relevanz, Umsetzbarkeit und Wirkung mit einer Punkteskala von 1 bis 10.»
Aufgaben aufteilen	Zerlegung komplexer Aufgaben in kleinere Schritte, die nacheinander bearbeitet werden können, um bessere Ergebnisse zu erzielen.	«Erstelle zuerst eine Liste potenzieller Aktivitäten für Jugendliche, und schätze anschliessend die erwartenden Kosten ein.»
Kontext	Bereitstellung relevanter Hintergrundinformationen, um die Aufgabe zu präzisieren.	«Die Aktivität soll in ländlichem Gebiet stattfinden, wo es bislang noch keine gleichwertigen Programme gab.»
Zielgruppe	Zielgruppe des generierten Texts, einschliesslich gewünschtem Schwierigkeitsgrad oder Stil.	«Erkläre das Konzept so, dass Eltern von betroffenen Kindern sowie Jugendbetreuer es leicht verstehen können.»
Ton	Tonalität oder Sprachstil des generierten Texts, angepasst an die Zielgruppe oder den Zweck.	«Verwende einen einfühlsamen und ermutigenden Ton, der Eltern Mut macht, an solchen Programmen teilzunehmen.»
Halluzination	Strategie zur Eindämmung falscher oder erfundener Antworten.	«Antworte nur, wenn du sicher bist. Wenn du es nicht weisst, frage nach oder weise darauf hin.»
Daten	Relevante Daten, die dem Modell übergeben werden, um den Kontext zu bereichern oder die Aufgabe zu präzisieren.	«Nutze die folgende Journalinformationen und erstelle einen Bericht in den Kategorien: Schule, Alltag und Gesundheit.»

Tabelle 2: Methoden für optimales Prompting [30, Kap. Overview of the Five Principles of Prompting ff.], [16, Kap. The Potential Complexity of a Prompt].

Die Interaktion mit einem LLM kann mehrere Iterationen umfassen, um möglichst präzise und ergänzende Antworten zu erhalten. Die nachfolgende Grafik zeigt, wie sich die verschiedenen Komponenten eines Prompts iterativ anpassen lassen, um die Qualität der Ergebnisse zu verbessern. Jede Iteration erlaubt eine gezielte Verfeinerung, indem fehlende Details ergänzt oder spezifische Anforderungen präzisiert werden. Dieser iterative Ansatz ermöglicht, Schritt für Schritt auf ein genaues und verwertbares Ergebnis hinzuarbeiten [16, Kap. The Potential Complexity of a Prompt].

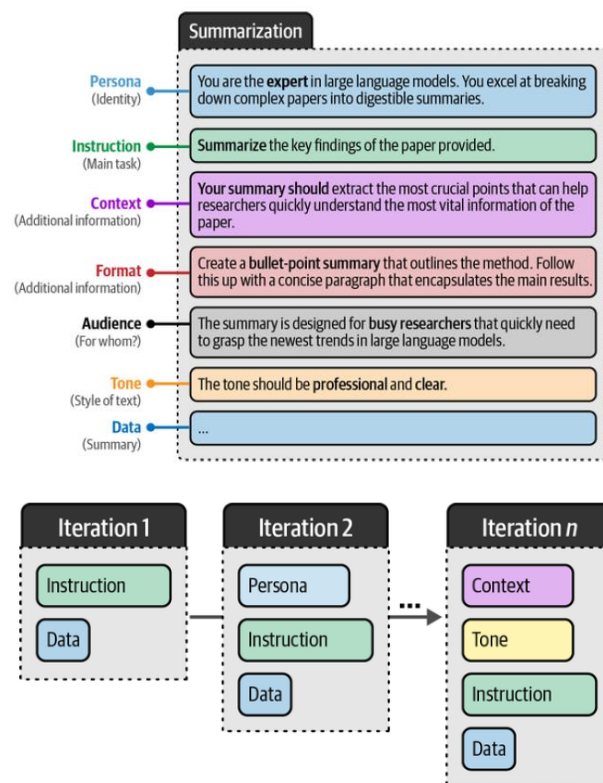


Abbildung 23: Anwendung von mehreren Iterationen beim Prompting [16, Kap. The Basic Ingredients of a Prompt].

Die Abbildung zeigt, wie verschiedene Elemente eines Prompts – von Instruktionen über den Kontext bis hin zu den Daten – iterativ verfeinert werden können, um die Qualität der generierten Ergebnisse zu optimieren. Dabei zeigt sich, dass die Qualität der übergebenen Daten eine ebenso zentrale Rolle spielt wie die Präzision der Anweisungen selbst.

Auch Stoica et al. haben in ihrer Studie die Wichtigkeit präziser Prompts untersucht. Mit der untenstehenden Grafik verdeutlichen die Autoren, wie ein unklarer Prompt zu unterschiedlichen Antworten führen kann. Die Modelle liefern auf die Frage, welches die schnellste Maschine ist, einerseits Informationen zu Computern und Fahrzeugen und andererseits einen Verweis auf den Teilchenbeschleuniger des CERN. In der Studie wird verdeutlicht, wie wichtig präzise Prompts in der Interaktion mit LLMs sind, insbesondere um Fehlfunktionen – wie Halluzinationen oder fehlerhafte Antworten – zu minimieren [31, S. 12].

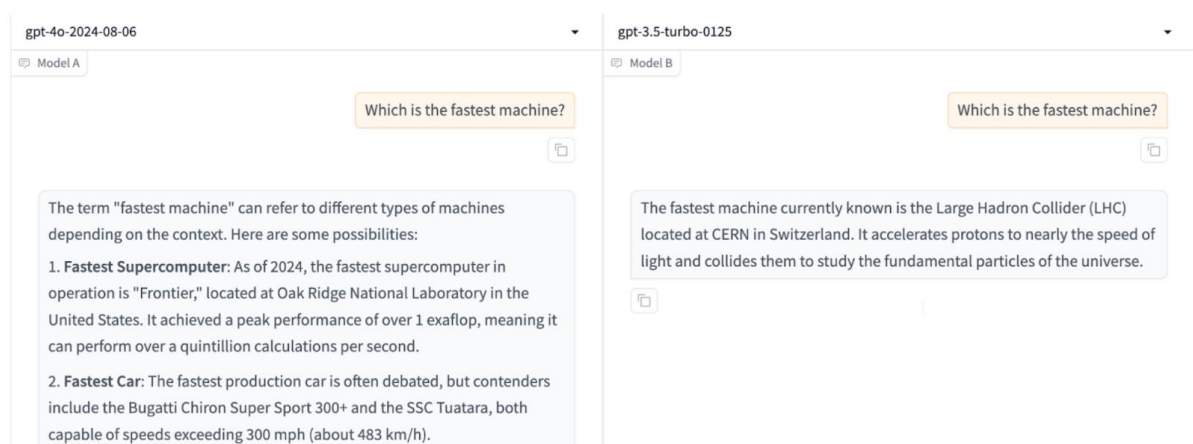


Abbildung 24: Mehrdeutige Prompts liefern unterschiedliche Resultate [31, S. 12].

Die Studie zeigt zudem, dass die Qualität eines Prompts nicht isoliert betrachtet werden kann, sondern direkt von der Klarheit und Präzision der zugrunde liegenden Spezifikationen abhängt. Besonders in juristischen und medizinischen Anwendungen sind klare Vorgaben essenziell, da fehlerhafte Prompts zu falschen, rechtlich problematischen oder sogar gefährlichen Antworten führen können. Stoica et al. schlagen daher vor, die Aufgabenstellung fundiert zu klären und präzise zu spezifizieren, um daraus bessere und gezieltere Prompts ableiten zu können [31, S. 3].

Diese Erkenntnis bildet den Ausgangspunkt für das nächste Kapitel, das untersucht, wie die Qualität und Verarbeitung der im Prompt übergebenen Daten die Funktionalität und Verlässlichkeit eines Modells beeinflussen. Hochwertige Prompts allein genügen nicht, wenn die zugrunde liegenden Daten fehlerhaft, unvollständig oder veraltet sind. Die Qualität der Eingabedaten ist daher ein entscheidender Faktor, der systematisch analysiert und bewertet werden muss, um sowohl die Effizienz des Modells als auch die Aussagekraft der generierten Ergebnisse zu gewährleisten.

4.2.4 Qualität der Eingabedaten

Weber bestätigt, dass nicht nur die Qualität der Daten im Modelltraining ausschlaggebend ist, sondern auch die Qualität der für Vorhersagen genutzten Eingabedaten. Selbst bei einem gut trainierten Modell können unzureichende oder unpassend strukturierte Eingangsdaten die Vorhersagen erheblich beeinträchtigen. Entscheidend ist, dass die Eingangsdaten syntaktisch und semantisch vergleichbar mit den ursprünglichen Trainingsdaten sind und denselben Verarbeitungsprozessen unterliegen. Ein zentrales Problem, das in diesem Zusammenhang auftreten kann, ist der sogenannte *Dataset Shift*. Dieses Phänomen beschreibt eine Veränderung in der Verteilung der Daten zwischen der Trainings- und der Anwendungsphase eines Modells. Wenn sich etwa die zugrunde liegende Struktur der Daten durch äussere Faktoren wie neue Marktbedingungen (z. B. Fachkonzepte), technologische Entwicklungen oder Änderungen im Nutzerverhalten verändert, kann dies dazu führen, dass ein Modell, das ursprünglich zuverlässige Vorhersagen lieferte, zunehmend ungenau wird. *Dataset Shift* verdeutlicht, dass die Gültigkeit eines Modells nicht allein durch die Qualität der Trainingsdaten garantiert werden kann, sondern kontinuierlich geprüft werden muss, ob die Eingabedaten den Trainingsdaten in ihrer Struktur, Semantik und Verarbeitung entsprechen. Dieser Ansatz unterstreicht, dass Datenqualität eine multidimensionale Herausforderung darstellt, die sowohl die Erstellung als auch die Nutzung von Modellen betrifft. Die Akzeptanz und Verlässlichkeit der Ergebnisse hängen wesentlich davon ab, wie sorgfältig die Daten in beiden Phasen geprüft und verarbeitet wurden [32, S. 246 ff.].

Auch Jones thematisiert die Herausforderungen, die mit schlechter Datenqualität einhergehen. Er betont, dass die Qualität der Daten der entscheidende Faktor für den Erfolg von KI-Modellen ist. Selbst ein einfaches Modell mit hochwertiger Datenbasis übertrifft ein komplexes Modell, das auf minderwertigen Daten aufbaut [33, Kap. Unlocking AI's potential with data]. Er bemerkt weiter, dass in datengetriebenen Organisationen häufig eine Diskrepanz zwischen den Erwartungen an die Datenqualität und den tatsächlichen Anforderungen besteht. Dies geschieht insbesondere dann, wenn die Erwartungen an die Datenqualität (z. B. von Prozessverantwortlichen oder der Leitung) nicht mit den Anforderungen der Datenproduzierenden (z. B. Fachpersonen der Sozialen Arbeit) abgestimmt sind. In solchen Fällen neigen Nutzerinnen und Nutzer dazu, unrealistisch hohe Annahmen über die Datenqualität zu treffen, was häufig zu einem Vertrauensverlust führt, wenn die Erwartungen nicht erfüllt werden. Um die Datenqualität bewerten zu können, ist es wichtig, klare Erwartungen festzulegen und die Daten anhand spezifischer Kriterien zu überwachen. Die folgende Tabelle stellt zentrale Aspekte der Datenqualität dar, die dabei helfen können, die Anforderungen an einen Datenbestand strukturiert zu erfassen und zu überprüfen [33, Kap. Assessing data quality]:

Dimension	Beschreibung	Kriterium für Bewertung
Vollständigkeit	Überprüfung, ob alle relevanten Daten vorhanden sind.	Sind alle notwendigen Informationen (z. B. Kategorien wie Pflege oder Verhalten) vollständig erfasst?
Eindeutigkeit	Sicherstellung, dass keine doppelten oder mehrdeutigen Einträge existieren.	Sind die Daten frei von Duplikaten oder widersprüchlichen Einträgen?
Zeitgerechtigkeit	Bewertung, ob die Daten aktuell und rechtzeitig erfasst wurden.	Sind die Daten zeitnah und repräsentieren sie die aktuelle Situation?
Gültigkeit	Überprüfung, ob die Daten den definierten Regeln und Standards entsprechen.	Entsprechen die Daten den formalen Vorgaben (z. B. konsistente Abkürzungen)?
Genauigkeit	Bewertung, ob die Daten die Realität korrekt widerspiegeln.	Sind die Daten präzise, frei von inhaltlichen Fehlern und korrekt in ihrer Zuordnung und Beschreibung?
Konsistenz	Sicherstellung, dass die Daten innerhalb und zwischen Quellen einheitlich sind.	Sind die Daten konsistent über verschiedene Einträge hinweg?
Nützlichkeit	Bewertung des intrinsischen Werts der Daten.	Tragen die Daten einen Mehrwert für den Anwendungsfall bei?

Tabelle 3: Dimensionen für die Bewertung von Datenqualität [33, Kap. Assessing data quality].

Eine inhaltliche Prüfung zeigt, dass die von Jones dargestellten Dimensionen auch bei Weber [32, S. 244 ff.] sowie in den Beschreibungen der *Data Management Association (DAMA)* [34] zu finden sind, was auf eine inhaltliche Übereinstimmung und damit auf ihre Validität hinweist.

Neben den in der Tabelle dargestellten Dimensionen der Datenqualität spielen auch weitere Aspekte eine wichtige Rolle. So kann laut Gunasekaran beispielsweise eine Sentimentanalyse [35] genutzt werden, um Emotionen und Meinungen in Texten automatisch zu erkennen und in Kategorien wie positiv, negativ oder neutral einzuteilen. Dadurch lassen sich Berichte oder andere schriftliche Quellen systematisch auswerten und gezielt für spezifische Anwendungsfälle – wie beispielsweise in der Sozialen Arbeit – nutzbar machen. Ergänzend dazu ermöglicht die Themenmodellierung die Identifikation wiederkehrender Muster und zentraler Themen in Textdaten, wodurch deren inhaltliche Struktur besser verstanden werden kann [36, S. 1].

Die Qualität der Eingabedaten und die präzise Gestaltung von Prompts sind entscheidende Faktoren für die Leistung eines Modells. Dennoch stellt sich die Frage, wie die generierten Texte hinsichtlich Genauigkeit, Relevanz und Kohärenz bewertet werden können. Das folgende Kapitel beleuchtet verschiedene Evaluierungsmethoden und zeigt auf, wie sowohl automatische als auch menschliche Bewertungsverfahren zur Beurteilung der Modellleistung eingesetzt werden.

4.2.5 Evaluierungsmethoden

Zur Bewertung automatisch generierter Texte haben sich zwei Methoden durchgesetzt: automatische Metriken und menschliche Evaluierung. Während automatische Metriken auf mathematischen Modellen basieren, um die Übereinstimmung eines generierten Textes mit einem Referenztext zu messen, umfasst die menschliche Evaluierung eine inhaltliche Analyse und Beurteilung durch (Fach-)Personen.

Schmidtová et al. analysierten in ihrer Studie, welche Methoden in wissenschaftlichen Arbeiten zur Qualitätssicherung von Texten verwendet werden. Dabei untersuchten sie 110 Beiträge von zwei wichtigen Konferenzen: der *International Conference on Natural Language Generation (INLG)* und der Jahrestagung der *Association for Computational Linguistics (ACL)*. Die Ergebnisse zeigen, dass automatische Metriken die dominierende Methode zur Textbewertung darstellen, während menschliche Evaluierungen seltener zum Einsatz kommen. In vielen Fällen wurden beide Ansätze kombiniert, um eine möglichst präzise Bewertung zu gewährleisten [37, S. 3]. Schmidtová et al. analysierten die automatisierten Methoden und ordneten diese zehn Kategorien zu. Die nachstehende Grafik zeigt die Verteilung in Bezug auf verschiedene Aufgaben. Die Prozentwerte über den Balken geben den Anteil der

jeweiligen Metrik-Familie innerhalb einer bestimmten Aufgabe an, während die absoluten Häufigkeiten unter den Prozentwerten angegeben sind [37, S. 7].

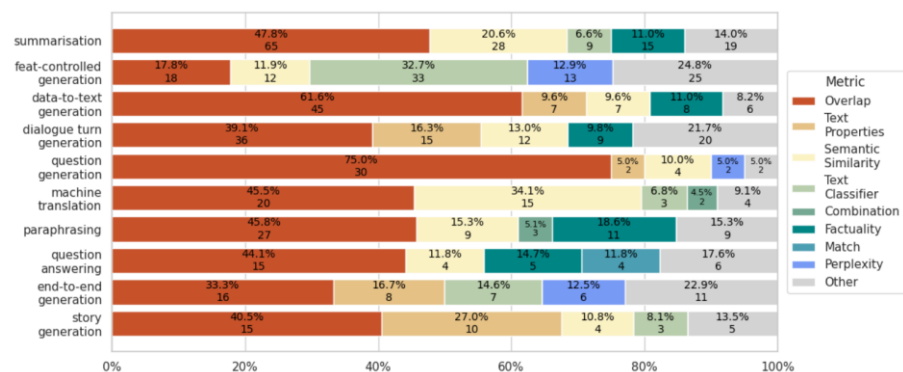


Abbildung 25: Verwendete Metrik-Familien für unterschiedliche NLP-Aufgaben im Jahr 2023 [37, S. 7].

Auffallend dabei ist der grosse Anteil der Metrik-Familie *Overlap*. Nach Schmidová et al. werden vor allem die beiden Metriken *BLEU* und *ROUGE* am häufigsten verwendet [37, S. 4]. Die Autoren weisen darauf hin, dass die Gründe für die Wahl dieser Evaluierungsmethoden oft unklar bleiben. Mehr als 80% der Arbeiten begründen die Wahl dieser Methoden ausschliesslich damit, dass sie bereits in vorhergehenden Studien verwendet wurden [37, S. 8]. Diese Praxis ist problematisch, da die bekannten Schwächen dieser Metriken nicht ausreichend berücksichtigt werden. Reiter kritisiert beispielsweise, dass *BLEU* nicht immer zuverlässige Ergebnisse liefert, da es häufig für Aufgaben genutzt wird, für die es ursprünglich nicht entwickelt wurde. Er argumentiert, dass *BLEU* nicht als primäre Bewertungsmethode eingesetzt werden sollte, insbesondere da methodische Verzerrungen bestehen, die schwer nachzuvollziehen sind. Zudem spiegelt die Methode nicht zuverlässig die Qualität von Texten in realen Anwendungskontexten wider. Dadurch sind sie insbesondere für sensible Bereiche wie die medizinische Forschung ungeeignet, in denen eine exakte und verlässliche Bewertung erforderlich ist [38, S. 7 f.].

Der Autor dieser Masterarbeit leitet daraus ab, dass Bewertungsmetriken vor ihrer Anwendung einer kritischen Überprüfung unterzogen werden müssen. Es ist sicherzustellen, dass automatisierte Evaluationsmethoden für die Textanalyse optimiert sind und den fachlichen Anforderungen genügen – insbesondere bei Texten mit hohen Anforderungen an Präzision und Verlässlichkeit. Zudem wird die Hypothese aufgestellt, dass eine menschliche Evaluierung zumindest unterstützend notwendig ist und in bestimmten Fällen unverzichtbar sein kann. Diese Annahme wird durch die Aussagen von van der Lee et al. gestützt, die die menschliche Bewertung von Ergebnissen als essenziell betrachten und empfehlen, sie häufiger und systematischer durchzuführen. Sie stellen fest, dass menschliche Bewertungen trotz verbesserter automatischer Messverfahren unerlässlich bleiben, um die Metriken zu überprüfen und ihre Nützlichkeit in verschiedenen Anwendungsbereichen sicherzustellen [39, S. 18]. Alamar und Grootendorst gehen noch einen Schritt weiter und bezeichnen die menschliche Evaluierung als *Goldstandard* der Textbewertung [16, Kap. Human Evaluation].

Neben der Wahl geeigneter Bewertungsmethoden spielt auch die klare Definition der zu erwartenden Ausgabe eine entscheidende Rolle. Stoica et al. argumentieren, dass eine präzise Spezifikation der gewünschten Ausgabe bereits vor der Generierung erfolgen sollte, um eine gezielte Bewertung zu ermöglichen. Zudem schlagen sie vor, dass LLMs ihre eigenen Ausgaben anhand vordefinierter Kriterien überprüfen können, sofern das Ausgabeformat und die erwarteten Inhalte eindeutig spezifiziert sind [31, S. 15]. Die Autoren heben hervor, dass Evaluierungsmethoden nicht isoliert betrachtet werden dürfen, sondern im Zusammenspiel mit den zugrunde liegenden Spezifikationen optimiert werden müssen [31, S. 3].

Ein alternativer Ansatz zur Verbesserung der Ergebnisqualität liegt jedoch bereits in der Modellarchitektur selbst. Multi-Agenten-Systeme ermöglichen eine effizientere Verarbeitung, indem mehrere spezialisierte KI-Agenten eine Aufgabe gemeinsam lösen. Das folgende Kapitel zeigt, wie dieser Ansatz insbesondere bei der automatisierten Textzusammenfassung zu präziseren Ergebnissen führt.

4.2.6 Agent AI

Ein vielversprechender Ansatz zur Verbesserung der Qualität KI-generierter Textzusammenfassungen ist der Einsatz von *Agent AI*. Ein KI-Agent ist ein autonomes System, das relevante Daten analysiert, Entscheidungen trifft und Aktionen ausführt, um ein bestimmtes Ziel zu erreichen. In Multi-Agenten-Systemen arbeiten mehrere dieser Agenten zusammen, indem sie Informationen austauschen und ihre Entscheidungen aufeinander abstimmen. Celikyilmaz et al. demonstrieren, dass dieser Ansatz besonders bei der automatisierten Textzusammenfassung vorteilhaft ist. Ihr Modell verteilt die Aufgabe der Textkodierung auf mehrere spezialisierte Agenten, die jeweils unterschiedliche Abschnitte eines Dokuments verarbeiten und sich dabei gegenseitig mit relevanten Informationen versorgen. Dies ermöglicht eine effizientere Extraktion relevanter Inhalte aus langen Texten und führt zu präziseren Zusammenfassungen. Die Studie zeigt, dass Multi-Agenten-Modelle mit interner Kommunikation eine höhere Qualität liefern als klassische Modelle mit nur einem einzigen Encoder [40, S. 1 f.]. Dieser Vorteil spiegelt sich insbesondere in höheren *ROUGE-Scores* wider. Aber auch in der manuellen Evaluierung erzielten die Zusammenfassungen der kommunizierenden Agenten durchgehend bessere Ergebnisse [40, S. 7].

Während Multi-Agenten-Systeme dazu beitragen können, die Qualität KI-generierter Texte zu verbessern, bleibt eine zentrale Herausforderung bestehen: die Nachvollziehbarkeit der Entscheidungen von KI-Modellen. Besonders bei komplexen, tiefen neuronalen Netzen sind die internen Entscheidungsprozesse oft nicht transparent. Genau hier setzt das Konzept der Explainable AI (XAI) an, das sich mit Methoden zur Erklärbarkeit von KI-Systemen befasst.

4.2.7 Explainable AI

KI wird zunehmend in Bereichen eingesetzt, in denen Entscheidungen unmittelbare Auswirkungen auf Menschen haben, etwa im Gesundheitswesen, der Justiz oder in der Sozialen Arbeit. Regulatorische Vorgaben, wie der *EU AI Act*, adressieren diese Herausforderung und weisen Herstellerinnen und Hersteller an, die Transparenz von KI-Modellen zu verbessern [8, S. 3]. Guidotti et al. beschreiben, dass viele Modelle, insbesondere tiefe neuronale Netze, als Black Boxes gelten, da ihre Entscheidungsprozesse für Menschen nicht nachvollziehbar sind. Dies kann problematisch sein, wenn es um Transparenz, Fairness und Verantwortlichkeit geht. Die Autoren betonen, dass Nutzerinnen bzw. Nutzer verstehen müssen, warum ein System eine bestimmte Entscheidung trifft, um Verzerrungen oder Fehleinschätzungen zu erkennen. Ohne Erklärbarkeit bleibt die Entscheidungsfindung in KI-Systemen intransparent und schwer überprüfbar [41, S. 2]. Um Black-Box-Modelle verständlicher zu machen, gibt es verschiedene Methoden. Manche Modelle sind von Natur aus interpretierbar, wie Entscheidungsbäume oder lineare Modelle. Andere benötigen nachträgliche Erklärungen, um ihre Funktionsweise sichtbar zu machen. Hierzu gehören Feature-Importance-Methoden, die zeigen, welche Eingaben besonders einflussreich waren, sowie regelbasierte Verfahren, die Entscheidungsregeln aus komplexen Modellen ableiten. Diese Methoden helfen, KI-Entscheidungen transparenter zu machen, sind aber nicht immer perfekt, da sie oft nur eine Annäherung an die tatsächliche Funktionsweise des Modells bieten. Trotz vieler Fortschritte gibt es weiterhin Herausforderungen. Oft besteht ein Spannungsfeld zwischen Genauigkeit und Interpretierbarkeit, da erklärable Modelle tendenziell weniger leistungsfähig sind als hochkomplexe neuronale Netze. Zudem fehlt eine einheitliche Metrik zur Bewertung der Qualität von Erklärungen [41, S. 34 ff.].

4.3 Implementierung

Die erfolgreiche Implementierung von KI-Modellen erfordert nicht nur eine theoretische Auseinandersetzung mit den Themen Datenqualität, Prompting und Evaluierungsmethoden, sondern auch eine sorgfältige Planung der technischen Infrastruktur. Dabei spielen sowohl Hardwareanforderungen als auch geeignete Implementationsverfahren eine zentrale Rolle, um eine effiziente und zuverlässige Nutzung der Modelle zu gewährleisten. Neben der technischen Machbarkeit gewinnt auch der Aspekt der Nachhaltigkeit zunehmend an Bedeutung. Der hohe Rechenaufwand moderner KI-Modelle führt zu einem erheblichen Energieverbrauch, weshalb Strategien zur Reduktion des ökologischen Fussabdrucks berücksichtigt werden müssen.

4.3.1 Hardwareanforderungen

Die benötigte Hardware kann erst dann präzise dimensioniert werden, wenn sowohl das zugrunde liegende Modell als auch die spezifischen Anforderungen der Aufgabe definiert sind. Erst auf dieser Basis lassen sich Faktoren wie Speicherbedarf, Rechenleistung und Skalierbarkeit bestimmen. Für den Betrieb und das Training grosser Sprachmodelle spielen mehrere zentrale Faktoren eine Rolle. Dazu zählen die Architektur des Modells, die Anzahl der Parameter, die Fenstergrösse (*Context Window*) und spezifische Implementierungstechniken. Die Fenstergrösse bestimmt beispielsweise, wie viele Tokens das Modell gleichzeitig verarbeiten kann. Ein grösseres Fenster ermöglicht es, längere Textabschnitte oder komplexe Zusammenhänge effizient zu erfassen, erhöht jedoch auch den Speicher- und Rechenaufwand, da mehr Informationen gleichzeitig analysiert und verarbeitet werden müssen. Die Grösse der Modelle wirkt sich signifikant auf die erforderliche Hardware aus, wobei leistungsstarke GPUs mit hohem Speicher oft unverzichtbar sind [16, Kap. Limited Resources Are All You Need]. Ein weiterer entscheidender Aspekt ist die Aufmerksamkeitsschicht (*Attention Layer*), die in der Transformer-Architektur den grössten Rechenaufwand verursacht. Sie wirkt sich daher sowohl auf die Leistungsfähigkeit als auch auf die Hardwareanforderungen eines Modells aus [16, Kap. Recent Improvements to the Transformer Architecture].

4.3.2 Implementationsverfahren

Alammar und Grootendorst beschreiben vor allem zwei Ansätze zur Implementierung grosser LLMs, welche nachfolgend grafisch dargestellt werden. Zum einen bieten verschiedene Firmen Schnittstellen (API) an, über die ihre Modelle auf eigener Hardware betrieben werden. Dies hat den Vorteil, dass die potenziell hohen Kosten für Betrieb, Wartung und Unterhalt der Infrastruktur vollständig von den Anbietenden übernommen werden. Zum anderen ist jedoch die Kontrolle über Datenflüsse sowie die Transparenz der Prozesse oft eingeschränkt und nur schwer nachvollziehbar [16, Kap. Interfacing with Large Language Models]. In der Regel werden verschiedene Kostenstrukturen angeboten, die sich durch Unterschiede in Funktionalität, Modellgrösse und Performance auszeichnen. So bietet *OpenAI* derzeit vier Optionen an, die es ermöglichen, mit unterschiedlichen Sprachmodellen und Variationen in Funktionalität und Leistung zu interagieren [42].

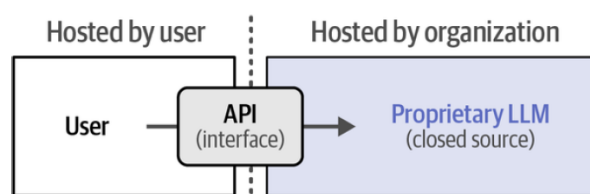


Abbildung 26: Anbindung von Closed Source LLMs über eine API [16, Kap. Proprietary, Private Models].

Aufgrund der fehlenden Kontrolle über die Datenflüsse und der unklaren Datentransparenz bei Implementierungen ist es notwendig, dieses Verfahren eingehend im Hinblick auf Datenschutz zu prüfen. Insbesondere da die sensiblen Daten den Einflussbereich der *GLAUX GROUP* verlassen, besteht das Risiko, dass personenbezogene oder sensible Informationen nicht ausreichend geschützt sind. Ab der

Schnittstelle (API) bleibt der weitere Prozess für Nutzerinnen und Nutzer jedoch intransparent, da die Daten extern verarbeitet werden. Diese Implementationsvariante scheint nur mit einer validen Datenanonymisierung realisierbar zu sein.

Eine Alternative dazu ist die Nutzung von Open-Source-Modellen auf eigener Infrastruktur. Diese ermöglicht es, die Kontrolle über die Daten vollständig zu behalten und die Prozesse transparent zu gestalten. Dieser Ansatz bietet volle Datenkontrolle, erfordert jedoch hohe Investitionen in Hardware und qualifiziertes Personal für Betrieb und Wartung. Alamar und Grootendorst weisen darauf hin, dass die technische Infrastruktur leistungsstark genug sein muss, um moderne Sprachmodelle effizient zu betreiben. Dabei sind robuste Hardware, wie GPUs mit hohem VRAM, und skalierbare Speicherlösungen essenziell [16, Kap. Proprietary, Private Models].

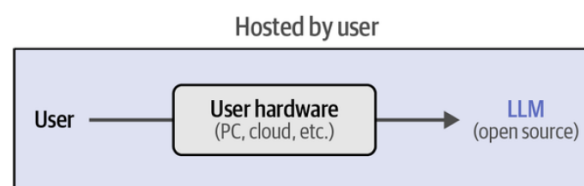


Abbildung 27: Implementierung von Open Source LLMs auf der eigenen Umgebung [16, Kap. Open Models].

Falls die Implementierung auf einer eigenen Infrastruktur in Betracht gezogen wird, sind zusätzliche Klärungen in den Bereichen Betrieb, Monitoring und Wartung nötig. Inhaltlich wird auf diese Punkte während der Masterarbeit jedoch nicht näher eingegangen.

4.3.3 Nachhaltigkeit

Wie bereits erwähnt, bieten vor allem grosse und leistungsfähige KI-Modelle den grössten Nutzen. Gleichzeitig bringen sie erhebliche Anforderungen an Rechenressourcen mit sich, was direkte Auswirkungen auf die ökologische Nachhaltigkeit hat. Sowohl der Energieverbrauch als auch der Einsatz von Rohstoffen müssen bei der Produktentwicklung berücksichtigt werden. Obwohl diese Masterarbeit das Thema nicht vertiefen kann, wird auf die Relevanz von Nachhaltigkeit hingewiesen, insbesondere durch die ISO 14'001-Zertifizierung der *GLAUX GROUP* und die damit verbundenen Verpflichtungen.

Schuh und Dölle stellen fest, dass die ökologische Nachhaltigkeit eine entscheidende Rolle bei der Entwicklung und Nutzung moderner Technologien spielt. Ziel ist, durch die Anwendung von Lean-Prinzipien – einem Ansatz zur Effizienzsteigerung durch die Vermeidung von Verschwendung – entlang der gesamten Wertschöpfungskette Ressourcen zu sparen und den Ressourceneinsatz bereits in der frühen Phase der Produktentwicklung zu minimieren. Dies umfasst unnötige Produkte oder Varianten nicht zu entwickeln, sondern den Fokus auf effiziente und zielgerichtete Lösungen zu legen [43, S. 11].

4.4 Zusammenfassung

Die technologische Entwicklung in der KI hat insbesondere in der Verarbeitung natürlicher Sprache bedeutende Fortschritte erzielt. Methoden wie maschinelles Lernen, Transformer-Modelle und RAG eröffnen neue Möglichkeiten, grosse Textmengen zu analysieren und Berichte automatisiert zu erstellen. Die Verarbeitung von Texten hat sich durch neuronale Netzwerke, Embeddings und moderne Architekturen wie Transformer grundlegend verändert und ermöglicht heute präzisere sowie effizientere Anwendungen. Trotz der Potenziale bestehen weiterhin technische Herausforderungen, darunter die Sicherstellung einer validen Datenqualität, die Vermeidung von Verzerrungen und die Einhaltung der Datenschutzkonformität. Auch die Wahl der optimalen Implementierungstechnik erfordert eine sorgfältige Abwägung zwischen Kosten, Kontrolle und Transparenz. Ein weiterer zentraler Punkt in der aktuellen Forschung ist die Notwendigkeit klarer Spezifikationen für die Entwicklung und Bewertung von LLM-basierten Systemen. Stoica et al. argumentieren, dass die KI-Entwicklung erst dann eine ingenieurmässige Disziplin wird, wenn sie durch überprüfbare Standards und präzise Aufgaben- sowie Lösungsspezifikationen geregelt ist [31, S. 18].

Diese Erkenntnisse verdeutlichen, dass KI-gestützte Berichterstellung zwar technisch realisierbar ist, jedoch eine klare strategische und methodische Ausrichtung erfordert. Die Implementierung erfordert nicht nur leistungsfähige Technologien, sondern auch eine enge Verzahnung mit den spezifischen Anforderungen der Sozialen Arbeit, um die Qualität, Ethik und Datenschutzkonformität sicherzustellen. Zudem ist die aktuelle Forschung von hoher Dynamik und schnellen Innovationszyklen geprägt, so dass kontinuierliche Weiterentwicklungen und Anpassungen notwendig sind.

4.4.1 Erkenntnisse zur Zielsetzung

Die Zielsetzung dieser Arbeit, KI zur automatisierten Berichterstellung in der Sozialen Arbeit zu untersuchen, findet in den technologischen Fortschritten eine solide Grundlage. Insbesondere Transformer-Modelle als zentrale Architektur des NLP bieten das Potenzial, Textdaten effizient zusammenzufassen und so die Dokumentationsprozesse zu optimieren. Die Implementierung solcher Technologien erfordert jedoch eine durchdachte Planung, insbesondere im Hinblick auf Datenschutz und Datenqualität. Die Erkenntnisse aus den vorangehenden Kapiteln verdeutlichen, dass Ansätze wie RAG helfen können, sensible Daten vor direkter Integration ins Modell zu schützen, während gleichzeitig relevante Informationen kontextbezogen verarbeitet werden. Dennoch ist die erfolgreiche Umsetzung solcher Systeme von der Verfügbarkeit hochwertiger und strukturierter Daten, der Einhaltung datenschutzrechtlicher Anforderungen sowie einer engen Zusammenarbeit zwischen technischen und sozialen Fachkräften abhängig. Der aktuelle Stand der Forschung zeigt, dass KI-basierte Technologien nicht nur die Effizienz steigern, sondern auch die Qualität und Einheitlichkeit der Ergebnisse verbessern können – vorausgesetzt, ethische und technische Standards werden berücksichtigt.

4.4.2 Erkenntnisse zu den Forschungsfragen

Welche technischen Anforderungen müssen erfüllt sein, um eine KI-basierte Berichterstellung in *socialweb* zu implementieren und gleichzeitig die datenschutzrechtlichen Vorgaben zu wahren?

Alammar und Grootendorst beschreiben verschiedene Methoden zur Integration von KI. Für Kontrolle und Transparenz wird eine lokale Implementierung der Modelle empfohlen. Sie weisen jedoch darauf hin, dass die technische Infrastruktur leistungsstark genug sein muss, um moderne Sprachmodelle effizient zu betreiben [16, Kap. Proprietary, Private Models]. Weiter beschreiben die Autoren, dass ein Ansatz wie RAG besonders geeignet ist, um datenschutzrechtliche Vorgaben einzuhalten. Er ermöglicht es, sensible Daten extern zu halten, ohne sie in das Modell zu integrieren, und optimiert zugleich die Modellabfragen [16, Kap. From Search to RAG].

Eine detaillierte Betrachtung dieser Aspekte findet sich in den Kapiteln der untenstehenden Tabelle.

Themen	Relevante Kapitel
Technische Infrastruktur	4.3.1 Hardwareanforderungen
Modellimplementierung	4.3.2 Implementationsverfahren
Datenschutz, Transparenz und Kontrolle	4.1.8 Retrieval Augmented Generation, 4.2.5 Evaluierungsmethoden, 4.2.6 Agent AI, 4.2.7 Explainable AI

Tabelle 4: Themen und relevante Kapitel für die Beantwortung der 1. technischen Forschungsfrage.

Welche methodischen Anpassungen in der Datenerfassung und Datenaufbereitung werden benötigt?

Die Qualität der Eingabedaten ist entscheidend für die Genauigkeit von KI-Modellen. Weber betont, dass die Daten strukturiert und mit den Trainingsdaten vergleichbar sein müssen, um Verzerrungen zu vermeiden [32, S. 246 ff.]. Jones beschreibt, dass eine Standardisierung der Datenerfassung essenziell ist, um Fehler zu minimieren und die Vergleichbarkeit sicherzustellen. Dazu gehören einheitliche Erfassungsrichtlinien, feste Datenformate und strukturierte Eingabemasken [33, Kap. Assessing data quality].

Themen	Relevante Kapitel
Datenqualität und Vergleichbarkeit	4.2.2 Trainingsdaten und ihre Implikationen
Standardisierung der Datenerfassung	4.2.4 Qualität der Eingabedaten
Vermeidung von Verzerrungen	4.2.2 Trainingsdaten und ihre Implikationen, 4.2.3 Prompting, 4.2.4 Qualität der Eingabedaten, 4.2.5 Evaluierungsmethoden, 4.2.6 Agent AI, 4.2.7 Explainable AI

Tabelle 5: Themen und relevante Kapitel für die Beantwortung der 2. technischen Forschungsfrage.

Ist eine Datenanonymisierung zwingend erforderlich, um KI datenschutzkonform in *socialweb* zu integrieren?

Alammar und Grootendorst beschreiben, dass Datenanonymisierung essenziell für eine datenschutzkonforme Implementierung ist, wenn die Modelle nicht auf eigener Hardware betrieben werden. Wird das Modell hingegen auf einer dedizierten Hardware betrieben, ist eine Anonymisierung nicht zwingend erforderlich, sofern sichergestellt wird, dass alle Daten lokal verarbeitet werden [16, Kap. Interfacing with Large Language Models]. Ansätze wie RAG ermöglichen es zudem, sensible Informationen extern zu speichern und dennoch kontextspezifisch für KI-Anwendungen nutzbar zu machen, ohne die Privatsphäre zu gefährden. [16, Kap. From Search to RAG].

Themen	Relevante Kapitel
Datenanonymisierung	4.1.8 Retrieval Augmented Generation, 4.3.2 Implementationsverfahren
Lokale vs. externe Verarbeitung	4.3.2 Implementationsverfahren
Datenschutzkonforme KI-Nutzung	4.1.8 Retrieval Augmented Generation, 4.2.7 Explainable AI

Tabelle 6: Themen und relevante Kapitel für die Beantwortung der 3. technischen Forschungsfrage.

Wie lässt sich die Ergebnisqualität der KI-generierten Berichte bewerten und kann diese durch Techniken wie Prompting optimiert werden?

Die Ergebnisqualität KI-generierter Berichte kann sowohl automatisch als auch manuell bewertet werden. Reiter kritisiert jedoch, dass Metriken wie *BLEU* nicht immer zuverlässige Ergebnisse liefern, da sie oft für Aufgaben genutzt werden, für die sie ursprünglich nicht entwickelt wurden [38, S. 7-8]. Van der Lee et al. betonen ebenfalls, dass menschliche Bewertungen unerlässlich sind, um automatische Metriken zu validieren und deren Anwendbarkeit zu überprüfen [39, S. 18]. Zur Optimierung der Ergebnisqualität zeigen Alammar und Grootendorst, dass Prompting eine effektive Technik ist, um KI-Modelle gezielt auf spezifische Aufgaben auszurichten und deren Genauigkeit zu steigern. Sie heben hervor, dass Prompting ein iterativer Prozess ist, der eine schrittweise Anpassung der Eingabe erfordert, um optimale Ergebnisse zu erzielen [16, Kap. The Potential Complexity of a Prompt].

Themen	Relevante Kapitel
Automatische vs. menschliche Bewertung	4.2.5 Evaluierungsmethoden
Zuverlässigkeit automatischer Metriken	4.2.5 Evaluierungsmethoden
Bedeutung menschlicher Validierung	4.2.5 Evaluierungsmethoden
Optimierung durch Prompting	4.2.3 Prompting

Tabelle 7: Themen und relevante Kapitel für die Beantwortung der 4. technischen Forschungsfrage.

Wie entstehen Verzerrungen (Bias) und mit welchen Methoden lassen sich diese minimieren?

Verzerrungen (Bias) in KI-Modellen können aus unterschiedlichen Gründen vorkommen. Laut Strasser und Niedermayer können unausgewogene oder einseitige Daten bestehende Vorurteile verstärken, während veraltete Informationen die Relevanz der Ergebnisse beeinträchtigen [3, S. 121 ff.]. Kreissl und von Laufenberg bestätigen, dass Modelle zuverlässig Verzerrungen reproduzieren, wenn in den Trainingsdaten Stereotypen und kulturelle Vorurteile enthalten sind [13, S. 240]. Dies zeigt sich laut Rigotti et al. auch in bekannten Fällen wie dem diskriminierenden Rekrutierungstool von Amazon [26]. Stoica et al. betonen weiter, dass Verzerrungen in LLMs häufig auf unklare oder unvollständige Spezifikationen sowie die damit verbundenen Prompts zurückzuführen sind. Eine präzisere Definition der Aufgabenstellung und Mechanismen zur iterativen Überprüfung der Modellantworten könnten dazu beitragen, Bias systematisch zu minimieren [31, S. 3]. Zur Minimierung von Bias sind mehrere Methoden erforderlich: Erstens empfehlen Alammari und Grootendorst den Einsatz mehrsprachiger Modelle, um sprachliche Verzerrungen zu reduzieren [16, Kap. Token Embeddings]. Zweitens verbessern Open-Source-Modelle wie *BLOOM* die Transparenz, da ihre Trainingsdaten einsehbar sind, während Closed-Source-Modelle wie *GPT-4* weniger nachvollziehbar sind [16, Kap. Token Embeddings]. Drittens können laut IBM regelmäßige Aktualisierungen der Trainingsdaten helfen, veraltete oder unausgewogene Inhalte zu vermeiden [27].

Themen	Relevante Kapitel
Ursachen von Verzerrungen	4.2.2 Trainingsdaten und ihre Implikationen, 4.2.4 Qualität der Eingabedaten, 4.2.5 Evaluierungsmethoden
Einfluss von Stereotypen und Vorurteilen	4.2.2 Trainingsdaten und ihre Implikationen
Einfluss unklarer Spezifikationen und Prompts	4.2.3 Prompting, 4.2.5 Evaluierungsmethoden
Methoden zur Bias-Minimierung	4.2.2 Trainingsdaten und ihre Implikationen, 4.2.3 Prompting, 4.2.5 Evaluierungsmethoden, 4.2.7 Explainable AI
Mehrsprachige Modelle zur Reduktion sprachlicher Verzerrungen	4.2.2 Trainingsdaten und ihre Implikationen, 4.2.3 Prompting
Regelmäßige Aktualisierung der Trainingsdaten	4.2.2 Trainingsdaten und ihre Implikationen, 4.2.4 Qualität der Eingabedaten

Tabelle 8: Themen und relevante Kapitel für die Beantwortung der 5. technischen Forschungsfrage.

5 Stand der Forschung Teil 2 – Perspektiven der Sozialen Arbeit

5.1 Anwendungsmöglichkeiten von KI in der Sozialen Arbeit

Gemäss Linnemann et al. bietet KI zahlreiche Anwendungsmöglichkeiten in der Sozialen Arbeit, insbesondere im Bereich der Datenanalyse und als Unterstützung bei Entscheidungsprozessen. Ein prominentes Beispiel ist das *Predictive Risk Modelling*, das dazu beiträgt, potenzielle Gefährdungen wie Kindeswohlverletzungen frühzeitig zu identifizieren. Dabei werden umfangreiche Datenquellen algorithmisch ausgewertet, um Fachkräfte bei der Risikoabschätzung zu unterstützen, wie dies beispielsweise durch das *Allegheny Family Screening Tool* in den USA umgesetzt wird. Forschungsergebnisse deuten darauf hin, dass solche datenbasierten Modelle präzise Ergebnisse liefern können und in manchen Fällen herkömmlichen Methoden überlegen sind [6, S. 206 ff.].

Linnemann et al. beschreiben, dass auch die Nutzung von Chatbots und digitalen Assistenzsystemen Potenziale bieten, um Personen niedrigschwellige Zugänge zu sozialen Dienstleistungen zu ermöglichen. Diese Systeme können für die Erstberatung oder die Bereitstellung von Informationen eingesetzt werden und entlasten damit Fachkräfte in zeitintensiven Routineaufgaben. Besonders in Beratungssituationen, in denen eine sofortige persönliche Betreuung nicht möglich ist, stellen Chatbots eine wertvolle Ergänzung dar [6, S. 202 ff.]. Darüber hinaus zeigt der Einsatz von NLP grosse Potenziale in der Sozialen Arbeit. Methoden der Sprachverarbeitung ermöglichen es, grosse Mengen natürlicher Sprache, wie z. B. Gesprächsprotokolle oder Dokumentationen, effizient zu analysieren und zu strukturieren. Dies könnte besonders bei der Entwicklung von Entscheidungsunterstützungssystemen nützlich sein. Pilotprojekte, wie das *MAEWIN-Projekt* an der *FH Bielefeld*, haben gezeigt, dass NLP-basierte Systeme helfen können, Ähnlichkeiten zwischen Fällen aufzuzeigen und Fachkräfte bei der Entscheidungsfindung zu unterstützen. Linnemann et al. heben dabei hervor, dass solche Systeme eine gezielte Schulung der Fachkräfte erfordern, um eine reflektierte Anwendung und die richtige Interpretation der Ergebnisse zu gewährleisten [6, S. 5 ff.]. Ein weiteres Einsatzfeld von NLP liegt im Wissensmanagement. Im Projekt *Lernende Systeme in der Beratung* der *Caritas* wird NLP eingesetzt, um Beratungsprotokolle und Expertenprofile zu durchsuchen. Ziel ist es, Beratende schnell mit relevanten Informationen und erfahrenen Kolleginnen und Kollegen zu vernetzen. Dies könnte den Zugang zu Fachwissen verbessern und den Erfahrungsaustausch zwischen Fachkräften fördern. Dennoch bleibt abzuwarten, ob solche Systeme die erhofften Vorteile tatsächlich bieten, da bisher keine umfassenden Evaluierungen vorliegen [6, S. 7 f.]. Linnemann et al. benennen ein weiteres wichtiges Anwendungsfeld: Das automatisierte Erstellen von Berichten. Diese Systeme können effizient Gesprächsprotokolle oder Fallberichte generieren und so den Dokumentationsaufwand für Fachkräfte erheblich reduzieren. Dadurch haben Fachkräfte die Möglichkeit, mehr Zeit für die direkte Arbeit mit betreuten Personen aufzuwenden und gleichzeitig die Qualität der Dokumentation zu steigern. Sie weisen jedoch darauf hin, dass derzeit wissenschaftlich belastbare Ergebnisse solcher Systeme noch ausstehen [6, S. 12 ff.].

Abschliessend lässt sich festhalten, dass die Anwendungsmöglichkeiten von KI in der Sozialen Arbeit vielseitig sind. Sie reichen von der Risikoabschätzung über die Unterstützung bei Beratungen bis hin zur verbesserten Dokumentation und Wissensvernetzung. Dennoch bleiben die Weiterentwicklung und Evaluierung solcher Technologien entscheidend, um sicherzustellen, dass sie die Profession der Sozialen Arbeit unterstützen, ohne zentrale Werte und Prinzipien zu gefährden.

5.2 Verantwortung und Ethik in der Anwendung

Ein zentraler ethischer Aspekt der Digitalisierung in der Sozialen Arbeit ist die verantwortungsbewusste Reflexion. Dabei geht es darum, die Auswirkungen digitaler Technologien nicht nur auf die Klientinnen bzw. Klienten, sondern auch auf die Profession selbst zu berücksichtigen. Kaminsky beschreibt, dass eine solche Reflexion sowohl die positiven Potenziale als auch die möglichen negativen Folgen der Digitalisierung für die Professionalität und die Werte der Sozialen Arbeit abwägen muss. Dabei ist besonders hervorzuheben, dass KI zwar als unterstützendes Werkzeug dienen kann, die komplexe und kontextbezogene Urteilskraft von Fachpersonen jedoch niemals ersetzen darf. Soziale

Arbeit erfordert nicht nur technische Effizienz, sondern auch Empathie, ethische Abwägungen und die Fähigkeit zur individuellen Fallbearbeitung – Qualitäten, die KI-Systemen von Natur aus fehlen [44, S. 2 ff.].

Auch Gutwald et al. weisen auf die potenziellen Gefahren der De-Professionalisierung hin, die mit einer zunehmenden Abhängigkeit von Technologien einhergehen können. Forschungen zeigen, dass Fachpersonen in der Sozialen Arbeit befürchten, dass ihre Rolle durch digitale Systeme abgewertet oder ersetzt werden könnte. Dadurch würde die notwendige menschliche Nähe und Empathie in der Facharbeit beeinträchtigt. Weiter wird darauf verwiesen, dass digitale Lösungen im Bereich der Sozialen Arbeit einem höheren Massstab an ethischen und sozialen Anforderungen genügen müssen als in anderen Bereichen [45, S. 6]. Entsprechend wird betont, dass ihre Entwicklung und Implementierung nur durch eine enge Zusammenarbeit zwischen Fachpersonen der Sozialen Arbeit und Expertinnen bzw. Experten der Informatik ethisch vertretbar gestaltet werden kann [45, S. 3].

Görder beschreibt weitere ethische Herausforderungen und Verantwortlichkeiten, die mit dem Einsatz von KI in der Sozialen Arbeit einhergehen. Dabei hebt er hervor, dass KI-Systeme einerseits das Potenzial haben, Prozesse effizienter und präziser zu gestalten, andererseits jedoch auch erhebliche Risiken bergen, die kritisch reflektiert werden müssen. Ein zentrales Thema ist die Gefahr ungerechtfertigter Diskriminierung. Algorithmen basieren auf bestehenden Datensätzen, die oft gesellschaftliche Vorurteile enthalten. Diese können durch KI verstärkt werden, was besonders in sensiblen Bereichen wie der Sozialen Arbeit schwerwiegende Folgen haben kann. Es ist daher notwendig, Trainingsdaten und Entscheidungsmechanismen sorgfältig zu prüfen und deren Anwendung transparent zu gestalten [46, S. 5]. Darüber hinaus thematisiert Görder die Frage gerechtfertigter Diskriminierung. Er argumentiert, dass der Einsatz von KI in vielen Fällen nicht zwangsläufig gleiche Behandlungen für alle Personen anstreben muss, sondern unter bestimmten Bedingungen gezielte Differenzierungen gerechtfertigt sein können. Folgt man dem klassischen Prinzip der Gerechtigkeit *Gleiches gleich und Ungleiches ungleich zu behandeln*, kann eine unterschiedliche Behandlung ethisch vertretbar sein, wenn sie auf objektiven und fairen Kriterien basiert. So könnte KI beispielsweise Risikofaktoren identifizieren, die bislang übersehen wurden, und dadurch gezieltere Präventionsmassnahmen ermöglichen. Zudem verweist Görder auf die gesellschaftlichen Implikationen von KI. Digitale Systeme können bestehende soziale Ungleichheiten verschärfen und individuelles Verhalten beeinflussen, was weitreichende ethische Fragen aufwirft. Fachpersonen der Sozialen Arbeit sind daher gefordert, die Funktionsweise und die Auswirkungen dieser Technologien kritisch zu hinterfragen und aktiv an ihrer verantwortungsvollen Nutzung mitzuwirken. [46, S. 7]. Abschliessend wird die Notwendigkeit eines bewussten und verantwortungsvollen Umgangs mit KI in der Sozialen Arbeit betont. Er hebt die besondere Bedeutung von Transparenz und Nachvollziehbarkeit hervor, insbesondere in sensiblen Anwendungsbereichen. Die Gestaltung und Nutzung von KI sollten stets darauf abzielen, das Wohlergehen der Klientinnen und Klienten zu fördern und zugleich bestehende soziale Ungleichheiten nicht zu verstärken. Ein solcher Ansatz ist aus ethischer Sicht unverzichtbar [46, S. 9].

Der Einsatz digitaler Arbeitsmittel erfordert auch laut Kaminsky nicht nur technische Expertise, sondern auch die Fähigkeit, deren Auswirkungen auf Daten und deren Nutzung kritisch zu hinterfragen. Die Komplexität dieser Technologien zeigt sich daran, dass ihre Funktionsweise oft nur mit erheblichem Aufwand nachvollziehbar ist. Es ist nicht immer unmittelbar ersichtlich, wie digitale Mittel – sei es Software, Apps oder andere Werkzeuge – technisch arbeiten, welche Zugriffe auf Daten sie ermöglichen und welche Folgen diese Zugriffe haben könnten. Daher ist es besonders für Fachpersonen der Sozialen Arbeit essenziell, eine angemessene Folgenabschätzung vorzunehmen, um verantwortungsvolle Entscheidungen über den Einsatz solcher Mittel treffen zu können [44, S. 3]. Um diesen Anforderungen gerecht zu werden, bedarf es gezielter Schulungsmassnahmen, die sowohl auf die Medienbildung als auch auf die Förderung ethisch-moralischer Argumentationskompetenz abzielen. Nur so kann die Profession sicherstellen, dass ihre zentralen Werte und Ziele gewahrt bleiben und gleichzeitig die Potenziale der Digitalisierung verantwortungsvoll genutzt werden [44, S. 17]. Neben den technischen und ethischen Aspekten der Digitalisierung ist der Schutz sensibler Daten von entscheidender Bedeutung. Unbeabsichtigte Nebenfolgen, wie etwa Verstösse gegen Datenschutzrechte, müssen systematisch identifiziert und berücksichtigt werden. Dies erfordert eine sorgfältige Untersuchung und Berücksichtigung möglicher Risiken bei der Entscheidungsfindung. Die Einhaltung höchster Datenschutzstandards ist unerlässlich, um das Vertrauen der betreuten Personen in die Facharbeit und die verantwortungsvolle Nutzung digitaler Technologien zu gewährleisten [44, S. 3].

Neben den ethischen Fragestellungen spielt die Qualität der Dokumentation und Berichte eine entscheidende Rolle. Eine unzureichende Struktur oder uneinheitliche Datenbasis stellt dabei nicht nur ein technisches Problem dar, sondern beeinflusst auch die professionelle Verantwortung und Entscheidungsfindung.

5.3 Dokumentation und Berichtswesen

Eine professionelle Dokumentation und Berichte übernehmen in der Sozialen Arbeit laut Reichmann eine zentrale Rolle und bilden somit die Grundlage für fundierte Entscheidungen und den Nachweis von professioneller Handlungskompetenz: Sie speichern praxisrelevantes Wissen, kommunizieren Lösungsvorschläge und legen Handlungsansätze sowie Verantwortlichkeiten fest. Gleichzeitig dienen sie der Vorbereitung von Entscheidungen, der Legitimation von Vorgehensweisen und dem Nachweis von Wirkungen. Diese Texte sind in institutionelle Rahmenbedingungen eingebettet und müssen formale Vorgaben wie rechtliche Normen, institutionelle Regeln sowie methodische und berufsethische Standards erfüllen. Damit sind Berichte weit mehr als einfache Informationsdokumente – sie sind ein essenzielles Werkzeug zur Steuerung und Qualitätssicherung in der Sozialen Arbeit [47, S. 17].

Die Erstellung von Berichten stellt jedoch auch grosse Herausforderungen dar, da unstrukturierte oder intuitive Ansätze häufig zu mangelhaften Ergebnissen führen. Insbesondere spontanes Losschreiben kann zwar schnell zu umfangreichen Texten führen, doch mangelt es solchen Dokumenten oft an einer klaren Struktur und Konsistenz. Ohne eine systematische Herangehensweise besteht die Gefahr, dass wesentliche Inhalte unvollständig oder fehlerhaft dargestellt werden, was die Entscheidungsgrundlage erheblich beeinträchtigt. Die spätere Überarbeitung solcher Texte erfordert meist einen erheblichen Aufwand, sodass die anfänglichen Zeitgewinne wieder verloren gehen. Um qualitativ hochwertige Berichte zu erstellen, ist daher eine strukturierte Herangehensweise essenziell [47, S. 20]. Dies zeigt sich sowohl in der traditionellen Berichterstellung als auch in der Nutzung KI-basierter Systeme.

Institutionelle Vorgaben und rechtliche Rahmenbedingungen spielen bei der Erstellung von Berichten eine besonders wichtige Rolle. In der Jugendhilfe beispielsweise bereiten Fachkräfte gerichtliche Entscheidungen durch detaillierte Anträge und Einschätzungen vor. Dabei können diese Berichte rechtliche Konsequenzen, wie etwa den Entzug der elterlichen Sorge, unterstützen oder steuern. Die Qualität der Inhalte ist daher von grosser Bedeutung, da sie die Grundlage für sensible Entscheidungen bildet, die direkten Einfluss auf die Lebensrealität der Betroffenen haben. Dies verdeutlicht, wie wichtig klare Vorgaben und Standards bei der Berichterstellung sind, um eine konsistente und zielgerichtete Dokumentation zu gewährleisten und sich der damit verbundenen Machtausübung bewusst zu werden [47, S. 25]. Reinmann liefert eine Reihe von Vorschlägen und Textbeispielen, wie Dokumentation und das Berichtswesen implementiert werden können, betont jedoch dabei, dass es nicht möglich ist, auf allgemein gültige Vorlagen zurückzugreifen, da diese für den jeweiligen Einsatz und Arbeitsbereich angepasst werden müssen [47, S. 158].

Ein weiteres Problemfeld in der Dokumentation ist die Inkonsistenz innerhalb von Akten. Diese entstehen oft durch unterschiedliche Vermerke und Berichte, die von verschiedenen Fachkräften erstellt werden und sich hinsichtlich ihres Stils, Aufbaus und ihrer Qualität stark unterscheiden können. Darüber hinaus werden Widersprüche zwischen Dokumenten in der Praxis selten bereinigt und bleiben in der Akte bestehen. Auch arbeitsteilige Bearbeitungsprozesse und wechselnde Zuständigkeiten tragen dazu bei, dass Akten uneinheitlich und teils widersprüchlich wirken. Dies erschwert nicht nur die Nachvollziehbarkeit von Entscheidungen, sondern gefährdet auch deren Qualität. Um solche Inkonsistenzen zu vermeiden, ist eine systematische und standardisierte Dokumentation unabdingbar [47, S. 86].

Auch Mayring hat dieses Phänomen untersucht und stellt fest, dass institutionelle Vorgaben, wie ein einheitliches Kategoriensystem, ein fundamentaler Baustein für die systematische Auswertung von Daten ist. Die Vorgaben dienen nicht nur der Strukturierung und Standardisierung der Daten, sondern ermöglichen auch eine Nachvollziehbarkeit der Analyse durch verschiedene Beteiligte [48, S. 51]. In der Sozialen Arbeit wird oft interdisziplinär zusammengearbeitet. Mayring verdeutlicht mit der untenstehenden Grafik die Vielschichtigkeit der Datenauswertung, die als Basis für das Verfassen von

Automatisierte Berichterstellung mit Künstlicher Intelligenz, Version 2, 29.07.2025



Die Anwendung von Künstlicher Intelligenz in der Sozialen Arbeit bietet ein breites Spektrum an Möglichkeiten, wie beispielsweise die Unterstützung in der Risikoabschätzung, der Beratung oder der Berichterstellung. Besonders die Analyse von natürlichen Sprachdaten, die Verwendung von Chatbots und die Automatisierung von Berichtsprozessen heben sich als zentrale Anwendungsfelder hervor. Gleichzeitig sind diese Technologien mit ethischen Herausforderungen, datenschutzrechtlichen Anforderungen und der Notwendigkeit einer transparenten und reflektierten Implementierung verbunden. Dieses Kapitel zeigt, dass KI-gestützte Systeme insbesondere durch klare Vorgaben, hochwertige Datenquellen und strukturierte Prozesse erfolgreich in der Praxis der Sozialen Arbeit eingesetzt werden können. Dennoch bleibt die Frage, wie die Profession ihre Werte und Prinzipien wahren kann, zentral, da technologische Lösungen menschliche Empathie und Urteilsfähigkeit nicht ersetzen können.

5.4.1 Erkenntnisse zur Zielsetzung

Die Zielsetzung der Arbeit, die KI-gestützte Erstellung von Berichten in der Sozialen Arbeit zu untersuchen, steht im Einklang mit den aktuellen Forschungsperspektiven. Die Literatur zeigt, dass KI-basierte Systeme, wie NLP oder automatisierte Berichterstellung, das Potenzial haben, Fachkräfte bei der Dokumentation zu entlasten und die Qualität von Berichten zu verbessern. Dennoch wird deutlich, dass die erfolgreiche Implementierung solcher Technologien eine fundierte Datenbasis, klare institutionelle Standards und eine systematische Schulung der Fachkräfte erfordert. Die Erkenntnisse bestätigen, dass KI nicht nur Prozesse optimieren, sondern auch die professionelle Verantwortung und Qualität der Sozialen Arbeit stärken kann – vorausgesetzt, die ethischen und datenschutzrechtlichen Anforderungen werden erfüllt.

5.4.2 Erkenntnisse zu den Forschungsfragen

Die Analyse der Literatur zeigt, dass der Einsatz von KI in der Sozialen Arbeit vielfältige Möglichkeiten bietet, jedoch auch spezifische Herausforderungen mit sich bringt. Nachfolgend werden die Erkenntnisse zu den Forschungsfragen im Zusammenhang mit der Sozialen Arbeit zusammengefasst:

Wie und in welchen Bereichen wird KI bereits heute in der Sozialen Arbeit eingesetzt?

Gemäss Linnemann et al. wird KI vor allem in der Datenanalyse, Entscheidungsunterstützung und Dokumentation eingesetzt. Beispiele wie das *Predictive Risk Modelling*, etwa das *Allegheny Family Screening Tool* in den USA, zeigen, wie datenbasierte Ansätze Risiken wie Kindeswohlgefährdungen frühzeitig erkennen können [6, S. 206 ff.]. Weiter wird beschrieben, dass Chatbots den Zugang zu sozialen Dienstleistungen erleichtern und damit Fachkräfte bei Routineaufgaben entlasten können [6, S. 202 ff.]. Ein weiterer Bereich ist NLP, das bei der Analyse von Gesprächsprotokollen oder Fallakten hilft. Projekte wie *MAEWIN* zeigen, dass NLP-basierte Systeme Fachkräfte bei der Fallbearbeitung unterstützen können [6, S. 5 ff.]. Ebenso wird die automatisierte Berichterstellung als Potenzial gesehen, um den Dokumentationsaufwand zu reduzieren, auch wenn hierzu belastbare Ergebnisse noch ausstehen [6, S. 12 ff.].

Themen	Relevante Kapitel
Aktuelle Beispiele	5.1 Anwendungsmöglichkeiten von KI in der Sozialen Arbeit

Tabelle 9: Themen und relevante Kapitel für die Beantwortung der 1. Forschungsfrage im Zusammenhang mit der Sozialen Arbeit.

Welche ethischen Anforderungen und Verantwortlichkeiten müssen KI-Systeme in der Sozialen Arbeit erfüllen, um professionellen und sozialen Ansprüchen gerecht zu werden?

KI-Systeme in der Sozialen Arbeit müssen Transparenz, Fairness, Datenschutz und professionelle Verantwortung wahren. Kaminsky betont, dass KI Fachpersonen unterstützen, aber nicht ihre Urteilsfähigkeit ersetzen darf [44, S. 2 ff.]. Gutwald et al. warnen vor einer De-Professionalisierung, die Empathie und menschliche Nähe gefährden könnte [45, S. 3]. Görder weist darauf hin, dass KI bestehende Vorurteile verstärken kann, weshalb transparente Entscheidungsmechanismen essenziell sind. Gleichzeitig kann eine gezielte Ungleichbehandlung ethisch vertretbar sein, wenn sie objektiv begründbar ist [46, S. 7]. Kaminsky hebt zudem die Bedeutung des Datenschutzes hervor: Fachpersonen müssen digitale Systeme verstehen, um Risiken zu vermeiden und das Vertrauen der betreuten Personen zu erhalten [44, S. 17].

Themen	Relevante Kapitel
Professioneller Umgang	5.2 Verantwortung und Ethik in der Anwendung

Tabelle 10: Themen und relevante Kapitel für die Beantwortung der 2. Forschungsfrage im Zusammenhang mit der Sozialen Arbeit.

Welche spezifischen Herausforderungen und Anforderungen ergeben sich in der Sozialen Arbeit bei der Dokumentation und der Erstellung von Berichten?

Die Dokumentation und Berichterstellung in der Sozialen Arbeit ist zentral für fundierte Entscheidungen und rechtliche Nachweise. Reichmann betont, dass Berichte methodische und ethische Standards erfüllen müssen, während eine mangelnde Struktur oder uneinheitliche Datenbasis Fehler und Inkonsistenzen verursachen kann [47, S. 17]. Besonders in der Jugendhilfe sind präzise Berichte entscheidend, da sie gerichtliche Entscheidungen beeinflussen [47, S. 158]. Mayring rät zu einer strukturierten Herangehensweise, damit die Dokumentationsqualität verbessert werden kann. Er empfiehlt ein einheitliches Kategoriensystem, um Konsistenz und Vergleichbarkeit sicherzustellen. Sowohl manuelle als auch KI-gestützte Berichterstellung profitieren von einer gut strukturierten, qualitativ hochwertigen Datenbasis und klaren Standards [48, S. 51]. Zudem betont er, dass Dokumentation und Datenauswertung vielschichtige Prozesse sind, was er grafisch darstellt hat [48, S. 59].

Themen	Relevante Kapitel
Dokumentation und Erstellung von Berichten	5.3 Dokumentation und Berichtswesen

Tabelle 11: Themen und relevante Kapitel für die Beantwortung der 3. Forschungsfrage im Zusammenhang mit der Sozialen Arbeit.

6 Methoden Teil 1 – Technologische Umsetzung

Der Umgang mit den sensitiven Datenbeständen ist gesetzlich reguliert [9]. Basierend auf der technischen Klärung in den vorangegangenen Kapiteln wird eine Lösung auf der eigenen Infrastruktur bevorzugt, um die Einhaltung und Kontrolle der Bestimmungen nach ISO 27'001 sicherzustellen. Da jedoch unklar ist, ob die zeitlichen Ressourcen der Masterarbeit ausreichen, um eine eigene Umgebung mit entsprechend hoher Ergebnisqualität aufzubauen, werden mehrere Stossrichtungen definiert. Diese werden im Rahmen eines Proof of Concept (PoC) vorab evaluiert. Die Vorgehensweise folgt einem strukturierten Schema, das nachfolgend grafisch und textuell erläutert wird.

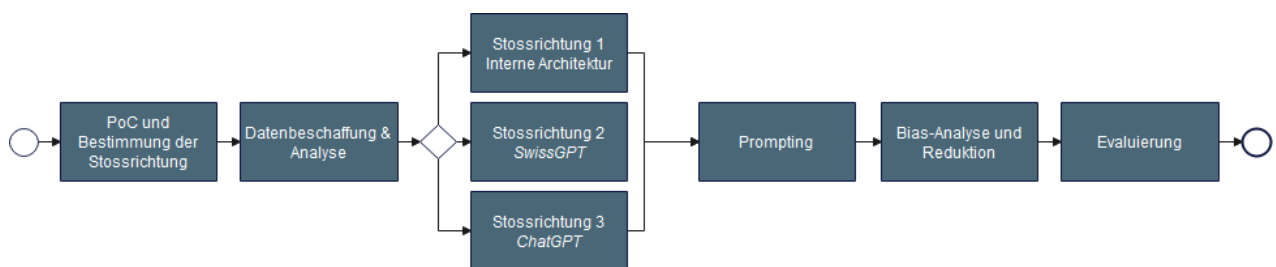


Abbildung 29: Methodik der Masterarbeit. Eigene Abbildung.

6.1 Proof of Concept und Bestimmung der Stossrichtung

Zur Klärung der technischen Möglichkeiten wird jeweils ein PoC erstellt. Anschliessend folgt die Auswertung und Auswahl einer Stossrichtung als Grundlage für die weitere Arbeit. Um einen direkten Vergleich der Stossrichtungen zu ermöglichen, sind klar definierte, praxisorientierte Kriterien festgelegt, die in der nachfolgenden Tabelle kurz beschrieben werden. Die Merkmale decken technische, rechtliche, ethische sowie wirtschaftliche Aspekte ab. Obwohl weitere ergänzt werden könnten, bieten die hier aufgeführten Kriterien eine praktische Grundlage für eine nachvollziehbare Bewertung:

Kriterium	Beschreibung	Skalierung	Faktor
Technische Aspekte			
Ergebnisqualität	Hier wird bewertet, wie präzise und zuverlässig die Ergebnisse sind. Qualität wird insbesondere anhand der Genauigkeit und Konsistenz der Ergebnisse gemessen.	1: Sehr schlecht – 5: Sehr gut	2
Systemintegration	Dieses Kriterium bezieht sich auf die technische Kompatibilität mit bestehenden Systemen und Prozessen. Eine einfache Integration minimiert Aufwand und Risiken.	1: Sehr schlecht – 5: Sehr gut	1
Usability und Community	Die Benutzerfreundlichkeit (Usability) sowie die Verfügbarkeit einer unterstützenden Community werden bewertet. Eine intuitive Bedienung und aktiver Support können entscheidend für den Erfolg sein.	1: Sehr schlecht – 5: Sehr gut	1
Robustheit	Dieses Merkmal beschreibt, wie stabil und fehlerfrei der PoC in unterschiedlichen Szenarien arbeitet. Ein robuster PoC zeigt reproduzierbare Ergebnisse und reagiert nicht empfindlich auf Variationen der Eingabedaten.	1: Sehr schlecht – 5: Sehr gut	2

Rechtliche und ethische Aspekte			
Datenschutz-konformität	Die Einhaltung gesetzlicher Datenschutzvorgaben, wie dem Schweizer Datenschutzgesetz (DSG) oder ISO 27'001, wird bewertet. Es wird eingeschätzt, inwieweit der PoC persönliche oder sensible Daten schützt und gesetzeskonform verarbeitet.	1: Sehr schlecht – 5: Sehr gut	2
Transparenz	Transparenz ist entscheidend, um Vertrauen aufzubauen und sicherzustellen, dass Nutzende und Verantwortliche die Arbeitsweise des Systems verstehen.	1: Sehr gering – 5: Sehr hoch	2
Wirtschaftliche Aspekte			
Aufwand für Entwicklung	Hier wird der zeitliche und technische Aufwand bewertet, der für die Erstellung des PoCs erforderlich ist. Ein geringerer Entwicklungsaufwand spricht für eine effizientere und praktikablere Lösung.	1: Sehr hoch – 5: Sehr gering	1.5
Aufwand für Betrieb und Wartung	Dieses Kriterium betrachtet die Ressourcen, die für den langfristigen Betrieb notwendig sind. Besonders wichtig sind dabei Aktualisierungen, Bugfixes und Kompatibilität.	1: Sehr hoch – 5: Sehr gering	1.5
Kosten für Nutzung	Dieses Kriterium bewertet die laufenden Kosten, die durch externe Anbieterinnen und Anbieter entstehen können. Beispiele sind Lizenzkosten, Abonnementgebühren oder kostenpflichtige Nutzungen, wie etwa eine Abrechnung pro Token.	1: Sehr hoch – 5: Sehr gering	1.5

Tabelle 12: Bewertungskriterien der Stossrichtung und Gewichtungen.

Für jede Stossrichtung wird ein eigener PoC erstellt, welcher jeweils diesem methodischen Schema folgt und sich durch die untenstehenden Inhalte unterscheidet:

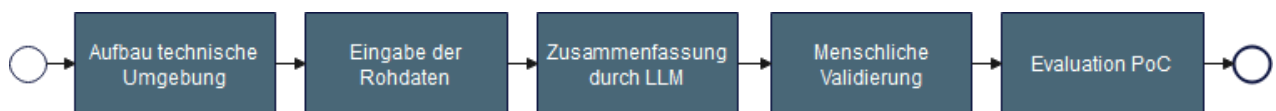


Abbildung 30: Methodik der PoCs. Eigene Abbildung.

Stossrichtung 1: Interne Architektur

Ertel beschreibt, dass sich Python für die Implementierung von KI gut eignet [12, S. 201]. Diese Einschätzung wird auch durch die *Computerwoche* geteilt [49]. Für Stossrichtung 1 wird daher eine eigene Python-Umgebung mit den entsprechenden Hilfsbibliotheken aufgebaut. Anschliessend soll mit unterschiedlichen Modellen ein allgemein bekannter Text zusammengefasst werden, um die Eignung des Ansatzes zu prüfen. Dies ermöglicht eine schnelle qualitative Einschätzung der Ergebnisse. Sollten die Modelle nicht in der Lage sein, diesen einfachen Text zufriedenstellend zusammenzufassen, würde dies frühzeitig auf grundlegende Defizite hinweisen. Dadurch kann der Aufwand für tiefergehende Tests eingespart werden. Selbstverständlich müsste anschliessend die Komplexität stufengerecht erhöht werden, bis fachlich fundierte und belastbare Resultate zu erwarten sind.

Stossrichtung 2: *SwissGPT* von *AlpineAI*

Am 7. November 2024 durfte der Autor den Fachworkshop *KI in Betrieben mit sozialem Auftrag – Herausforderungen und Nutzen* der INSOS Solothurn besuchen. INSOS ist der Branchenverband der Dienstleister für Menschen mit Behinderung [50]. Im Gespräch wurde *SwissGPT* der Firma *AlpineAI* als weitere Möglichkeit erwähnt, um KI sicher und insbesondere datenschutzkonform in Schweizer Produkten zu verwenden. In Stossrichtung 2 wird geprüft, ob *SwissGPT* über eine Schnittstelle in die Python-Umgebung von Stossrichtung 1 integriert werden kann, um rechenintensive Interaktionen auszulagern.

Stossrichtung 3: ChatGPT von OpenAI

Anfang 2023 machte *ChatGPT* Schlagzeilen, als der Chatbot eine Jura-Prüfung an der *Universität Minnesota* bestand [51]. Obwohl *GPT-4o* nicht das einzige populäre Sprachmodell ist [52, S. 15 ff.], wird es für die Stossrichtung 3 gezielt verwendet. In den letzten Monaten hat sich dieses Modell als robust und leicht zugänglich bewährt. Das Modell wird zudem von einer aktiven Community unterstützt und ist weit verbreitet. Der Text der Stossrichtung 1 wird durch *GPT-4o* von *OpenAI* zusammengefasst.

6.2 Datenbeschaffung, Analyse und Vorbereitung

Die Daten werden von der vordefinierten Institution unter Einhaltung der ISO-27'000-Richtlinien und unter Aufsicht des CISO beschafft und auf einer dedizierten Hardware gespeichert. Sie werden auf ihre Qualität hin geprüft, um zu beurteilen, ob sie für die Zusammenfassung und die gewählte Stossrichtung geeignet sind. Die allgemeinen Schritte der Datenvorbereitung umfassen die Datenbereinigung, Anonymisierung (je nach Stossrichtung) und Strukturierung, um eine effiziente Verarbeitung zu ermöglichen.

6.3 Umsetzung der Stossrichtung

Bei Stossrichtung 1: Interne Architektur

Die technischen Gegebenheiten von PoC 1 werden verfeinert und auf die sensiblen Datenbestände angewendet. Als Ergebnis wird ein zusammenfassender Bericht in deutscher Sprache erwartet.

Bei Stossrichtung 2: SwissGPT von AlpineAI

Die Datenbestände werden manuell anonymisiert, damit ein Rückschluss auf die Person nicht mehr möglich ist. Die technischen Gegebenheiten von PoC 1 werden verfeinert und auf die sensiblen Datenbestände angewendet. Als Ergebnis wird ein zusammenfassender Bericht in deutscher Sprache erwartet.

Bei Stossrichtung 3: ChatGPT von OpenAI

Die Datenbestände werden manuell anonymisiert, damit ein Rückschluss auf die Person nicht mehr möglich ist. Anschliessend werden pro Dossier Projekte in *ChatGPT* eröffnet und die anonymisierten Journale als Basis in den Dateien hinterlegt. Die Datenbestände werden anschliessend durch *GPT-4o* von *OpenAI* zusammengefasst. Als Ergebnis wird ein zusammenfassender Bericht in deutscher Sprache erwartet.

6.4 Prompting

Es wird untersucht, wie die Resultate durch Anpassung der Prompts verbessert werden können. Dies umfasst die Auswahl und Optimierung von Eingabeprompts, um die Qualität der generierten Ergebnisse zu maximieren. Dabei kommen die in Kapitel 4.2.3 beschriebenen Prompting-Techniken zur Anwendung. Die Effektivität der Prompts wird kontinuierlich bewertet und an die spezifischen Anforderungen der Anwendung angepasst.

6.5 Bias-Analyse und Reduktion

Im Rahmen der Bias-Analyse wird untersucht, ob Vorurteile (Bias) in den generierten Texten auftreten. Falls solche Verzerrungen identifiziert werden, werden verschiedene Techniken zur Reduktion eingesetzt (siehe Kapitel 4.2.3 *Prompting*), um die Objektivität und Fairness der Ergebnisse sicherzustellen.

6.6 Evaluierung

Wie bereits in Kapitel 4.2.5 *Evaluierungsmethoden* erläutert, ist eine menschliche Evaluierung – insbesondere für eine valide Beurteilung – unerlässlich. KI-generierte Inhalte müssen auf Kohärenz, inhaltliche Genauigkeit und mögliche Verzerrungen geprüft werden. Daher werden die generierten Berichte in dieser Arbeit manuell analysiert, um ihre Qualität, Konsistenz und Praxisrelevanz zu beurteilen.

7 Methoden Teil 2 – Wirtschaftliche Perspektiven

Neben einer fundierten technischen Betrachtung kommen ergänzend wirtschaftliche Methoden zum Einsatz, um die wirtschaftliche Tragfähigkeit und Risiken zu bewerten. Wie im Kapitel 3.4 *Abgrenzung* erwähnt, erfolgt keine tiefergehende Auseinandersetzung mit diesen Ansätzen und es wird auch keine Suche nach weiteren, potenziell moderneren Methoden unternommen. Es werden lediglich folgende Verfahren angewendet:

- Berechnung der Entwicklungs- und Betriebskosten
- Kosten-Nutzen-Analyse
- Sensitivitätsanalyse
- Risiko und Szenarioanalyse

8 Technische Ergebnisse

Um die optimale technische Umsetzung für diese Arbeit zu bestimmen, wurden drei unterschiedliche Stossrichtungen untersucht. Diese umfassen den Aufbau einer internen Architektur, den Einsatz von *SwissGPT* von *AlpineAI* sowie die Nutzung von *ChatGPT* von *OpenAI*. Ziel dieser Evaluierung war es, eine Lösung zu identifizieren, die sowohl technisch leistungsfähig als auch datenschutzkonform und wirtschaftlich tragfähig ist. Die folgenden Abschnitte präsentieren die Ergebnisse der einzelnen Stossrichtungen sowie die abschliessende Auswahl der am besten geeigneten Lösung für diese Arbeit.

8.1 Proof of Concept und Bestimmung der Stossrichtung

8.1.1 Stossrichtung 1: Interne Architektur

Es wurden verschiedene Modelle mit unterschiedlichen Methoden auf einer internen Umgebung getestet, wobei insbesondere das Buch *Natural Language Processing mit Transformatoren* von Tunstall et al. als praxisorientierte Referenz diente [53]. Die Bewertung erfolgte anhand vordefinierter Kriterien, die in der untenstehenden Tabelle dargestellt sind. Die getesteten Modelle erreichten weder die gewünschte Ergebnisqualität noch die angestrebte Robustheit. Alamar und Grootendorst weisen zudem darauf hin, dass eine Implementierung auf einer internen Architektur zu erhöhten Entwicklungs- und Betriebskosten führen würde, was sich negativ auf die wirtschaftliche Bewertung auswirkte [16, Kap. Open Models].

Kriterium	Beschreibung	Bewertung	Faktor	Produkt
Technische Aspekte				
Ergebnisqualität	Die Ergebnisqualität war unzureichend. Keines der eingesetzten Modelle war in der Lage, einen einfachen Text valide zusammenzufassen. Es fehlten wichtige Inhalte [54, L. 38], die Ausgabe enthielt Mischformen aus deutschen und englischen Wörtern [54, L. 71] oder es traten Wiederholungen einzelner Wörter auf [54, L. 46]. Zudem zeigte sich, dass die Modelle starke Halluzinationen generierten, wodurch der Kontext der Zusammenfassungen von der ursprünglichen Aufgabenstellung abwich [55, L. 10].	1	2	2
Systemintegration	Die Systemintegration konnte im PoC nicht abschliessend überprüft werden, da die Rahmenbedingungen innerhalb des Testszenarios begrenzt waren. Daher wird dieser Aspekt neutral bewertet.	3	1	3
Usability und Community	Es stehen zahlreiche Beispiele und Ressourcen, sowohl in Fachbüchern als auch im Internet, zur Verfügung, was diesen Punkt grundsätzlich positiv beeinflusst. Trotz der Verfügbarkeit dieser Ressourcen konnte die geforderte Ergebnisqualität nicht erreicht werden. Der Punkt wird dennoch hoch bewertet, da die Community eine wertvolle Unterstützung bietet. Es ist jedoch zu berücksichtigen, dass Entwicklerinnen und Entwickler fundierte Fachkenntnisse benötigen, um spezifische und valide Fragestellungen an die Community zu richten.	4	1	4
Robustheit des PoCs	Es zeigte sich, dass der PoC nicht robust war. Die zugrunde liegende Aufgabe wurde unterschätzt. Einerseits fehlte das notwendige Fachwissen, um die Ergebnisqualität zu verbessern,	1	2	2

	andererseits erwiesen sich die eingesetzten Methoden bereits bei einfachen Aufgaben als unzureichend.			
Rechtliche und ethische Aspekte				
Datenschutzkonformität	Die Datenschutzkonformität wird sehr hoch eingestuft, da die eingesetzten Modelle, Methoden und die Infrastruktur optimal auf die geltenden gesetzlichen Vorgaben abgestimmt werden können.	5	2	10
Transparenz	Die Transparenz wird ebenfalls hoch bewertet, da die Implementierung weitgehend gesteuert werden kann. Zum Einsatz kamen vortrainierte Modelle. Aufgrund der fehlenden Kontrolle über die Trainingsdaten wurde bei der Bewertung ein Punkt abgezogen.	4	2	8
Wirtschaftliche Aspekte				
Aufwand für Entwicklung	Der Entwicklungsaufwand wird als sehr hoch bewertet. Dies liegt einerseits an den eigentlichen Entwicklungsaufgaben und andererseits an der aufwendigen Validierung der Ergebnisse. Eine besondere Herausforderung besteht darin, dass fundiertes technisches KI-Wissen innerhalb der <i>GLAUX GROUP</i> erst noch aufgebaut werden muss.	1	1.5	1.5
Aufwand für Betrieb und Wartung	Der Betrieb erfordert den Aufbau einer eigenen Infrastruktur, deren Umfang im Rahmen des PoCs nur vage eingeschätzt werden konnte. Viele Modelle entwickeln sich rasant weiter, was einerseits Innovationen ermöglicht, andererseits jedoch auch zusätzliche Aufwände für Implementierung und Parametrisierung mit sich bringt. Diese Entwicklungen erfordern kontinuierliche Anpassungen und Wartung, um langfristig eine optimale Nutzung zu gewährleisten.	1	1.5	1.5
Kosten für Nutzung	Die Kosten für die Nutzung schneiden im Vergleich zu bestehenden Online-Angeboten voraussichtlich besser ab. Allerdings bleibt eine gewisse Unschärfe bestehen, da der Aufwand für Entwicklung und Betrieb nicht genau beziffert werden konnte. Aus diesem Grund wird dieser Punkt neutral bewertet.	3	1.5	4.5
Summe		-	-	36.5

Tabelle 13: Bewertung der Stossrichtung 1.

Die Gesamtsumme von 36.5 von möglichen 72.5 Punkten verdeutlicht, dass diese Stossrichtung wesentliche Herausforderungen mit sich bringt – insbesondere hinsichtlich Ergebnisqualität und Robustheit.

Ein besonders eindrückliches Beispiel zeigte sich beim Einsatz des Modells *google/pegasus-multi_news*, das die Zusammenfassung eines Grimm-Märchens mit einer frei erfundenen Biografie von Angela Merkel verknüpfte. Die generierte Ausgabe enthielt Behauptungen über Konzentrationslager, familiäre Verstrickungen in den Nationalsozialismus sowie mehrfach wiederholte, widersprüchliche Aussagen – alles ohne jeden Bezug zum Ausgangstext.

Dieses Beispiel verdeutlicht, dass generative Modelle nicht nur irrelevante, sondern auch potenziell rufschädigende Inhalte halluzinieren können. Der vollständige Modelloutput ist im Anhang 16.1 dokumentiert.

8.1.2 Stossrichtung 2: *SwissGPT* von *AlpineAI*

Bei Stossrichtung 2 wurde versucht, Informationen zu spezifischen technischen und datenschutzbezogenen Aspekten über *SwissGPT* zu erhalten. Leider konnte diese Stossrichtung aufgrund von Kommunikationsschwierigkeiten mit dem Anbieter nicht abschliessend evaluiert werden. Trotz wiederholtem Kontakt mit dem technischen Verantwortlichen (CTO) der Firma *AlpineAI* blieben mehrere Termin- und Informationsanfragen unbeantwortet. Da der Anbieter nicht reagierte, konnte kein PoC erstellt werden, wodurch eine abschliessende Bewertung nicht möglich war. Insgesamt verdeutlicht diese Erfahrung, dass eine enge Zusammenarbeit und klare Kommunikationswege zwischen Entwicklerinnen und Entwicklern sowie potenziellen Implementierungspartnerinnen und -partnern entscheidend sind, um eine zuverlässige Beurteilung neuer Technologien vornehmen zu können.

8.1.3 Stossrichtung 3: *ChatGPT* von *OpenAI*

Es wurden verschiedene Tests mit *ChatGPT (GPT-4o)* durchgeführt, um dessen Potenzial zu bewerten. Die Tests ergaben eine hohe Übereinstimmung der generierten Inhalte mit den Vorgaben und eine sehr einfache Systemintegration. Kritisch bleiben nach Alammar und Grootendorst rechtliche Aspekte wie der Datenschutz und die Transparenz, insbesondere bei sensiblen Daten, zu bewerten. [16, Kap. Proprietary, Private Models].

Kriterium	Beschreibung	Bewertung	Faktor	Produkt
Technische Aspekte				
Ergebnisqualität	Die Ergebnisqualität war hervorragend. <i>ChatGPT</i> lieferte schnelle, präzise und valide Resultate. Es wurden während des PoCs keine Halluzinationen oder Kontextfehler festgestellt [56, L. 5].	5	2	10
Systemintegration	Die Integration in bestehende Systeme ist sehr einfach. <i>ChatGPT</i> bietet umfangreiche Schnittstellen, die eine schnelle Einbindung in bestehende Prozesse ermöglichen [57].	5	1	5
Usability und Community	<i>ChatGPT</i> verfügt über eine grosse und aktive Community. Es gibt zahlreiche Ressourcen und Beispiele, die die Nutzung erleichtern. Dies erhöht die Benutzerfreundlichkeit erheblich [58].	5	1	5
Robustheit des PoCs	Der PoC zeigte sich als sehr robust und zuverlässig. Die Performance blieb stabil, auch wenn grössere Texte zusammengefasst wurden [59].	5	2	10
Rechtliche und ethische Aspekte				
Datenschutzkonformität	Da sensitive Daten für die Verarbeitung an <i>OpenAI</i> weitergeleitet werden müssen, ist der Datenschutz sehr kritisch zu bewerten. Die Daten müssen präzise anonymisiert und idealerweise zusätzlich manuell überprüft werden, was die Handhabung sehr erschweren würde.	1	2	2
Transparenz	Die Transparenz bleibt unklar, da die Trainingsdaten nicht vollständig nachvollziehbar sind. Es fehlen Details zu den zugrunde liegenden Modellen und Prozessen.	2	2	4
Wirtschaftliche Aspekte				
Aufwand für Entwicklung	Der Aufwand für die Entwicklung ist sehr gering, da <i>ChatGPT</i> sofort einsatzbereit ist und nur minimale Anpassungen benötigt.	5	1.5	7.5

Aufwand für Betrieb und Wartung	Der Aufwand für Betrieb und Wartung ist ebenfalls gering. Die Infrastruktur wird durch den Anbieter bereitgestellt, wodurch interne Aufwände minimiert werden.	5	1.5	7.5
Kosten für Nutzung	Die Kosten für die Nutzung sind hoch und werden durch die Anzahl Tokens beeinflusst. Um dies langfristig wirtschaftlich tragfähig zu gestalten, müsste ein dediziertes Geschäftsmodell erarbeitet werden, das den Bedarf der Institutionen berücksichtigt und eine bedarfsgerechte Abrechnung ermöglicht.	2	1.5	3
Summe		-	-	54

Tabelle 14: Bewertung Stossrichtung 3.

Mit einer Gesamtsumme von 54 von möglichen 72.5 Punkten zeigt die Stossrichtung 3 ein höheres Potenzial. Besonders hervorzuheben sind die hervorragende Ergebnisqualität, die einfache Integration und der geringe Entwicklungsaufwand. Kritisch bleiben jedoch gemäss Alammar und Grootendorst die Aspekte Datenschutz und Transparenz, die in sensiblen Einsatzbereichen nicht nur zusätzlichen Aufwand erfordern, sondern auch die Integration in bestehende Prozesse erheblich erschweren oder im Extremfall sogar unmöglich machen könnten [16, Kap. Proprietary, Private Models].

8.1.4 Evaluation und Auswahl Stossrichtung

Aufgrund der getesteten Kriterien wurde *ChatGPT* als praktikabelste Option für diese Arbeit eingestuft. Die getesteten Methoden der ersten Stossrichtung lieferten mangelhafte Ergebnisse hinsichtlich Qualität und Robustheit. Es wurde klar, dass das Fachwissen des Autors nicht ausreicht, um die technischen und methodischen Herausforderungen dieses Ansatzes zu bewältigen. Zeitgleich haben auch andere Teams in der *GLAUX GROUP* ähnliche Erfahrungen gemacht und bestätigt, dass die Komplexität hoch ist. Stossrichtung 2 konnte nicht abschliessend bewertet werden, da der Anbieter keine relevanten Informationen lieferte. Im Gegensatz dazu überzeugte *GPT-4o* durch einfache Implementierung, ausgezeichnete Ergebnisqualität und eine hohe Zuverlässigkeit. Die Nutzung von Schnittstellen ermöglichte eine rasche Integration und die Ergebnisse waren präzise und stabil, selbst bei komplexen Aufgaben. Diese Vorteile minimieren nicht nur den Entwicklungsaufwand, sondern gewährleisten auch eine effiziente Nutzung der verfügbaren Zeit. Datenschutzrechtliche Aspekte bleiben relevant, können jedoch durch eine frühzeitige Anonymisierung der Daten vor ihrer Verwendung in dieser Arbeit wirksam adressiert werden. Insgesamt bietet *ChatGPT* die optimale Kombination aus Qualität, Effizienz und Pragmatismus für die spezifischen Anforderungen dieser Arbeit.

8.2 Datenbeschaffung, Bereinigung und Anonymisierung

8.2.1 Datenbeschaffung

Die Beschaffung der Daten erfolgte unter strikter Einhaltung der geltenden ISO-Richtlinien, um höchste Standards im Bereich Sicherheit und Datenschutz zu gewährleisten. Hierfür wurde ein dedizierter virtueller Server bereitgestellt, der ausschliesslich für die Datenverarbeitung in diesem Projekt genutzt und nach Abschluss wieder abgebaut wurde. Die Bereitstellung des Servers ermöglichte eine abgeschottete Umgebung, die die Sicherheit der sensiblen Daten während des gesamten Prozesses gewährleistete. Der Beschaffungsprozess wurde in enger Zusammenarbeit mit dem *CISO* der *GLAUX GROUP* abgestimmt. So konnte sichergestellt werden, dass sowohl die technischen als auch die organisatorischen Anforderungen des Datenschutzes eingehalten wurden. Der *CISO* prüfte und genehmigte den gesamten Prozess, um potenzielle Risiken auszuschliessen und die Einhaltung aller relevanten Vorgaben zu garantieren.

Die Auswahl der spezifischen Datensätze erfolgte durch die Institution selbst. Deren Verantwortliche identifizierten die relevanten Daten und stimmten dem Austausch zu. Dadurch wurde sichergestellt, dass ausschliesslich relevante und zweckgebundene Daten bereitgestellt wurden, die den Zielen der

Masterarbeit entsprechen. Die Daten wurden schliesslich am 18. Dezember 2024 auf den bereitgestellten Server übertragen.

8.2.2 Bereinigung und Anonymisierung

Im Rahmen der manuellen Bereinigung und Anonymisierung wurden die Daten aufbereitet, um die Einhaltung der Datenschutzvorgaben sicherzustellen. Dieser Prozess erfolgte in mehreren Schritten, um sowohl die Relevanz der Daten für das Projekt als auch den Schutz der beteiligten Personen zu gewährleisten:

- Die Daten wurden in drei separate Dossiers aufgeteilt, um eine bessere Strukturierung und Organisation zu erreichen.
- Einträge mit hochsensiblen Informationen, die potenziell ein erhebliches Risiko für die Privatsphäre der betroffenen Personen darstellten, wurden gelöscht.
- Spalten ohne Relevanz für die Analyse wurden entfernt.
- Technische Metainformationen wie HTML-Tags (`<div>`, `</div>`, `<p>`, `</p>`, `<br>`) wurden aus allen Einträgen entfernt, um die Daten in ein einheitliches und lesbares Format zu überführen.

Ein zentraler Bestandteil der Anonymisierung war die Entfernung aller personenbezogenen Informationen. Der Autor hat die Akten manuell gesichtet und sämtliche sensiblen Inhalte anonymisiert. Die Namen der Dossiers sowie die Bezeichnungen von Mitarbeitenden und externen Fachpersonen wurden geändert. Zusätzlich wurde der Name der Institution durch ein Pseudonym ersetzt. Dieser umfassende Bereinigungs- und Anonymisierungsprozess stellte sicher, dass alle verarbeiteten Daten den datenschutzrechtlichen Vorgaben entsprechen und weder direkte noch indirekte Rückschlüsse auf beteiligte Personen oder Institutionen möglich sind. Falls in den nachfolgenden Kapiteln Namen genannt werden, sind diese ausschliesslich anonymisiert.

8.3 Datenanalyse

Die Datenanalyse erforderte einen erheblichen manuellen Aufwand, um relevante Informationen zu extrahieren. Daher wurden von den ausgewählten Dossiers jeweils nur die Daten eines Jahres berücksichtigt. Diese Entscheidung wurde durch die Beobachtung, dass sich die inhaltlichen Themen über die Monate hinweg kaum veränderten, zusätzlich gestützt. Eine detaillierte Darstellung der Analyse sowie ergänzende Grafiken sind im Anhang zu finden.

Um fundierte Erkenntnisse aus den aufbereiteten Daten zu gewinnen, wurde sowohl eine quantitative als auch eine qualitative Analyse durchgeführt. Im folgenden Abschnitt werden zunächst die quantitativen Ergebnisse vorgestellt, welche bei den statistischen Methoden zur Untersuchung der Daten eingesetzt wurden.

8.3.1 Quantitative Analyse

Die durchgeführten Analysen zeigen einige interessante Aspekte und bilden eine erste Grundlage für die Ableitung von Erkenntnissen und Empfehlungen. Die Ergebnisse werden in der nachstehenden Tabelle aufgeführt:

Thema	Bewertung
Anzahl Einträge, Lücken	Die Anzahl der Einträge variiert erheblich zwischen den Dossiers. Dossier <i>K-2021-00003</i> umfasst mehr als doppelt so viele Einträge wie Dossier <i>K-2021-00060</i>: • Auch wenn Ersteres einen grösseren Zeitraum abdeckt (4 Monate mehr, +13%), ist dies allein vermutlich keine abschliessende Erklärung. • Ein weiterer möglicher Einflussfaktor ist die deutlich höhere Anzahl an Verfasserinnen und Verfasser bei Dossier <i>K-2021-00003</i>.

	In der Regel erstellen wenige Personen die meisten Einträge. Dies deutet auf eine zentrale Datenerfassung durch Kernteam-Mitglieder oder spezifische Bezugspersonen hin. Bis auf einen Monat im Dossier <i>K-2021-00003</i> gibt es keine Lücken in den Daten.
Anzahl Zeichen pro Eintrag	Der Medianwert der Anzahl Zeichen pro Eintrag und pro Dossier liegt zwischen 253 und 318 Zeichen. Einzelne Einträge weichen signifikant vom Median ab, was auf unterschiedliche Dokumentationspraktiken oder Schreibstile hindeutet. Beispiele: - Mitarbeiterin bzw. Mitarbeiter 21: 1750 bei Dossier <i>K-2021-00079</i> - Mitarbeiterin bzw. Mitarbeiter 26: 847 bei Dossier <i>K-2021-00060</i>
Kategorien und Markierungen	Zwei Kategorien dominieren: <i>01 Beobachtung</i> und <i>02 Gesundheit</i> . Die Verteilung der Kategorien und Markierungen ist ungleichmässig. Besonders hervorzuheben sind die vielen markierten Einträge im Dossier <i>K-2021-00003</i> , bei den Kategorien: <i>6. Ernährung/Verdauung</i> und <i>8. Liebe/Sexualität</i> .
Sentimentanalyse	Mit Hilfe einer Sentimentanalyse wurde die Stimmung der Einträge bewertet. Die Werte liegen zwischen -1 (negativ) und +1 (positiv). Dabei fiel auf: - Fachpersonen dokumentieren unterschiedlich, einige mit besonders hohen oder niedrigen Sentimentwerten. - Zwei Modelle (<i>nlptown/bert-base-multilingual-uncased-sentiment</i> und <i>oliverguhr/german-sentiment-bert</i>) zeigten unterschiedliche Ergebnisse, was die Notwendigkeit einer Validierung unterstreicht.

Tabelle 15: Quantitative Bewertung der Daten.

Die Resultate der quantitativen Analyse bieten einige Rückschlüsse, sind aber ohne qualitative Untersuchung nur bedingt aussagekräftig. Die Variabilität der Daten könnte auf Faktoren wie Fallkomplexität, Teamzusammenarbeit oder institutionelle Vorgaben zurückzuführen sein. Es ist daher wichtig, keine voreiligen Schlüsse zu ziehen, da die erhobenen Daten erst im Kontext einer weiteren Analyse und unter Berücksichtigung zusätzlicher Einflussfaktoren vollständig interpretiert werden können.

8.3.2 Qualitative Analyse

Die qualitative Analyse der Dossiers erfolgt anhand der Dimensionen von Jones und *DAMA* (siehe Kapitel 4.2.4 *Qualität der Eingabedaten*) und wird in der nachfolgenden Tabelle dargestellt:

Dimension	Bewertung
Vollständigkeit	Die Dossiers zeichnen sich durch eine weitgehende Vollständigkeit aus, da die Einträge kontinuierlich und täglich im Modul <i>Journal</i> erfasst wurden. Die stichprobenartige Überprüfung zeigt jedoch, dass teilweise unklare Überschneidungen mit dem Modul <i>Pflege</i> vorliegen. Manche Informationen, wie Medikamentengaben oder Vitalwerte, werden sowohl im <i>Journal</i> als auch im Modul <i>Pflege</i> dokumentiert, was zu einer Doppelspurigkeit führen kann. Darüber hinaus konzentrieren sich die Einträge häufig auf bestimmte Themenbereiche, wie Gesundheit und Verhalten, während andere Lebensaspekte, wie soziale Interaktionen oder Freizeitaktivitäten, weniger prominent dokumentiert sind. Es bleibt unklar, ob diese Bereiche absichtlich

	ausgeklammert wurden oder schlicht nicht Teil der Dokumentationsroutine sind. Eine umfassendere Integration dieser Lebensbereiche würde das Bild vervollständigen und die Datenbasis erweitern.
Eindeutigkeit	Die Eindeutigkeit der Einträge ist überwiegend gegeben, da in den meisten Dossiers keine nennenswerten Duplikate (identische Einträge) oder widersprüchlichen Inhalte gefunden wurden. Im Dossier <i>K-2021-00003</i> wurden neun Duplikate identifiziert, die auf eine ineffiziente Erfassungspraxis hinweisen könnten. In Dossier <i>K-2021-00060</i> wurde ein Duplikat festgestellt. Die Formulierungen innerhalb der Einträge sind meist klar, es gibt jedoch gelegentlich Verwirrung durch uneindeutige Bezeichnungen. Beispielsweise werden Abkürzungen wie <i>nie</i> genutzt. Diese verweist auf eine Person, könnte aber durch eine KI auch als Verneinung interpretiert werden. Die eindeutige Zuordnung solcher Begriffe ist essenziell, um Missverständnisse zu vermeiden. Ausserdem erschwert die uneinheitliche Schreibweise von Namen die Verarbeitung und Anonymisierung der Daten.
Zeitgerechtigkeit	Die Erfassung der Daten erfolgt weitgehend zeitnah. Rückwirkende Einträge sind selten und betreffen in der Regel spezifische Ereignisse, die nachträglich ergänzt wurden. Diese Praxis scheint jedoch nicht systematisch zu sein, da die zeitliche Nähe zwischen dem Ereignis und der Erfassung oft eingehalten wird. Die zeitliche Konsistenz ist ein positiver Aspekt, da sie sicherstellt, dass die Einträge die aktuelle Situation widerspiegeln. Dennoch könnten klarere Richtlinien zur Nachbearbeitung von Einträgen sinnvoll sein, um die Kohärenz und Nachvollziehbarkeit auch bei verspäteten Dokumentationen sicherzustellen.
Gültigkeit	Die Gültigkeit der Daten wird durch die uneinheitliche Nutzung von Abkürzungen und Kategorien beeinträchtigt. Es gibt keinen einheitlichen Standard, der vorgibt, welche Begriffe und Abkürzungen genutzt werden sollen. Während einige Mitarbeitende systematisch normierte Abkürzungen einsetzen, wie sie beispielsweise im Dossier <i>K-2021-00079</i> vorkommen, fehlt in anderen Dossiers eine vergleichbare Konsistenz. Die Kategorien, wie <i>Beobachtung</i> oder <i>Gesundheit</i> werden oft breit genutzt, wodurch sich die Inhalte überschneiden und die Trennung von Themen erschwert wird. Beispielsweise werden pflegerische Themen wie Körperpflege und medizinische Massnahmen häufig in der Kategorie <i>Beobachtung</i> erfasst, was die Vergleichbarkeit und Auswertung der Daten beeinträchtigt. Standardisierte Vorgaben könnten die Gültigkeit der Einträge verbessern und die Daten für weiterführende Analysen nutzbarer machen.
Genauigkeit	Die Genauigkeit der Einträge ist durch eine Vielzahl von Rechtschreibfehlern, informellen Formulierungen und inkonsistenten Kategorien beeinträchtigt. Begriffe wie <i>brav</i> oder <i>vollgepinkelt</i> wirken unprofessionell und könnten durch neutralere Alternativen ersetzt werden. Darüber hinaus werden Beobachtungen häufig ohne konkrete Verweise oder Belege dokumentiert, wie im Fall von medizinischen Ergebnissen, die nicht weiterverfolgt wurden. Die Breite der Kategorien trägt zusätzlich zu Ungenauigkeiten bei. Unter <i>Beobachtung</i> finden sich beispielsweise Themen, die spezifischeren Kategorien zugeordnet werden könnten, etwa <i>Pflege</i> oder <i>Verhalten</i>. Diese Praxis erschwert eine präzise Analyse und Nutzung der Daten, da die Inhalte nicht eindeutig voneinander abgegrenzt sind.
Konsistenz	Die Konsistenz zwischen den Modulen <i>Journal</i> und <i>Pflege</i> wurde stichprobenartig überprüft und war in den meisten Fällen zufriedenstellend. Es gibt jedoch Fälle, in denen Informationen redundant oder widersprüchlich erfasst wurden. Beispielsweise werden pflegerische Massnahmen sowohl

	im <i>Journal</i> als auch im Modul <i>Pflege</i> dokumentiert, ohne dass klar ist, welches Modul die primäre Quelle darstellt. Innerhalb der Dossiers gibt es Diskrepanzen in der Sprache und Terminologie. Während in einigen Einträgen von <i>Klient</i> oder <i>Bewohner</i> die Rede ist, wird in anderen von <i>Kunde</i> gesprochen. Diese inkonsistente Verwendung erschwert die Vergleichbarkeit der Einträge und deutet auf fehlende Standards hin.
Nützlichkeit	Die Daten dokumentieren den Tagesverlauf der betreuten Personen und unterstützen das Team in der interdisziplinären Zusammenarbeit. Inwiefern sie sich für die automatisierte Erstellung von Berichten eignen, lässt sich ohne weiterführende Tests nur schwer einschätzen.

Tabelle 16: Qualitative Bewertung der Daten nach Jones und DAMA.

8.3.3 Zusammenfassung und Handlungsempfehlungen

Es wurde sowohl eine quantitative als auch eine qualitative Analyse durchgeführt, um fundierte Erkenntnisse zu gewinnen. Während die quantitative Analyse hilfreiche Leitlinien und erste Anhaltspunkte lieferte, zeigte die qualitative Analyse, dass die Dokumentation weitgehend vollständig ist, jedoch in bestimmten Bereichen Optimierungspotenzial besteht. Besonders die Standardisierung und die Einführung klarer institutioneller Vorgaben sind dabei von entscheidender Bedeutung. Basierend auf den Erkenntnissen werden im Folgenden konkrete Empfehlungen ausgesprochen:

- Die Einführung verbindlicher Abkürzungen und die Vermeidung von Dialekt und informeller Sprache würden die Konsistenz erheblich steigern.
- Sammelkategorien wie *Beobachtung* sollten durch spezifischere ersetzt werden.
- Markierungen für berichts- und alltagsrelevante Inhalte könnten die Qualität der Dokumentation erhöhen.
- Die Einhaltung von Datenschutzrichtlinien sollte konsequent sichergestellt werden.
- Missverständliche oder unangemessene Begriffe sollten vermieden werden.
- Regelmässige Schulungen und Feedbackmechanismen könnten die Dokumentationsqualität nachhaltig verbessern.

Es wird empfohlen, dass die Institution klar definiert, welchem Zweck die Dokumentation dient. Durch die Umsetzung der Empfehlungen kann die Dokumentation qualitativ und zukunftssicher gestaltet werden.

Hinweis für die weitere Bearbeitung

Bei der Prüfung der Datenqualität wurde festgestellt, dass es Überschneidungen zwischen den Modulen *Journal* und *Pflege* gibt. Eine weiterführende Analyse zeigte ausserdem, dass zahlreiche Medikamente abgegeben werden, deren Erwähnung in einem Bericht fachlich sinnvoll erscheint. Daher wurde beschlossen, den strukturierten Verlauf der Pflegedokumentation – bestehend aus Medikamentabgaben und Pflegemassnahmen – als zusätzliche Datengrundlage für die Berichterstellung einzubeziehen.

8.4 Erstellung von Berichten

8.4.1 Einfaches Prompting

Zunächst wurde *ChatGPT* mit einem einfachen Prompt aufgefordert, die vorhandenen Journaldaten in einem Bericht zusammenzufassen. Die ursprünglich gelieferten Dateien konnten nicht verarbeitet werden. Erst nach der Umwandlung in ein *CSV-Format* war das System in der Lage, die Informationen zu verarbeiten.

Der initiale Prompt lautete: «*Erstelle einen Bericht über Lukas Schmid*». Der generierte Bericht erwies sich jedoch als unstrukturiert und ohne erkennbaren logischen Zusammenhang. Im Kontext der

Sozialen Arbeit waren die Ergebnisse unzureichend. Der Bericht enthielt überwiegend quantitative Informationen, beispielsweise über Textelemente oder die Anzahl der Verfasserinnen bzw. Verfasser [60].

Dieses Verhalten wurde bei allen untersuchten Dossiers beobachtet. Die Ergebnisse machten deutlich, wie essenziell präzisere und ergänzende Prompts sind. Daher wurde der Ansatz iterativ verbessert, um präzisere Ergebnisse zu erzielen.

8.4.2 Ergänzendes und verbessertes Prompting

Ergänzende und verbesserte Prompts [61] führten dazu, dass die generierten Berichte eine hohe sprachliche Kohärenz aufwiesen. Die KI war in der Lage, Berichte über die hochgeladenen Dossiers in kurzer Zeit zu generieren und dabei eine fachlich korrekte sowie eloquente Sprache zu verwenden. Besonders positiv hervorzuheben ist, dass die generierten Berichte iterativ optimiert werden konnten. So liessen sie sich gezielt durch spezifische Prompts wie *«Erwähne im Abschnitt Wohnen den Klinikaufenthalt»* anpassen und nach Belieben ergänzen.

Jedoch traten auch erhebliche Schwächen auf. Die Antworten auf identische Prompts waren oft unterschiedlich, was die Reproduzierbarkeit und Transparenz der Ergebnisse erheblich einschränkte. Die generierten Berichte enthielten inhaltliche Abweichungen und Fehler. Die Analyse identifizierte mehrere Herausforderungen in der KI-basierten Berichterstellung. Es wurden Inhalte generiert, die nicht in den hochgeladenen Daten enthalten waren. Untenstehend eine Aufzählung weiterer und unterschiedlicher Herausforderungen aus den untersuchten Fällen:

- Obwohl dem Modell strukturierte Daten zum Thema *Pflege und Medikation* zur Verfügung gestellt wurden, wurde das Medikament *Melatonin 5 mg* aufgeführt, obwohl dieses in der Akte nicht erwähnt wurde. Auf Nachfrage teilt *ChatGPT* mit: *«Die Erwähnung von "Melatonin 5 mg" wurde entfernt, da es in den bereitgestellten Daten nicht vorkommt.»* [62, S. 11].
- Medikamente wurden teilweise nicht mehr mit dem richtigen Namen aufgeführt, sondern nur noch mit *Medikament A*, *Medikament B*. Nach entsprechender Aufforderung wurde dieser Umstand wieder bereinigt [63, S. 4].
- Im Dossier *K-2021-0003* schreibt die KI: *«In Krisensituationen wurde Valium 5mg verwendet, was laut Dokumentation zu einer schnellen Beruhigung führte»* [64, S. 4].

Ab Juli wurde die Medikation angepasst: *Valium* ersetzte die *CBD*-Behandlung. Diese Information wäre im Bericht erwähnenswert gewesen, ebenso die Frage, ob sich die Situation unter dem neuen Medikament verändert hat. Hinweis: Auch in den ursprünglichen Journaldaten wurde kein entsprechender Eintrag über eine Anpassung der Medikation gefunden.

- Zur Behandlung akuter und chronischer schizophrener Syndrome und bei akuten psychomotorischen Erregungszuständen wurde das Medikament *Haldol* abgegeben. Nach dem Klinikaufenthalt wurde das Medikament im April vermehrt verabreicht und auch über das Jahr war eine Steigerung der Medikation feststellbar (siehe Abbildung unten). Das Medikament wurde in den Berichten nicht explizit erwähnt, sondern durch allgemeine Beschreibungen ersetzt. Beispiel: *«Bedarfsmedikation: In Krisensituationen kam ein beruhigendes Medikament zum Einsatz. Hierbei wurde eine rasche Wirkung dokumentiert, ohne relevante Nebenwirkungen.»* [64, S. 4].

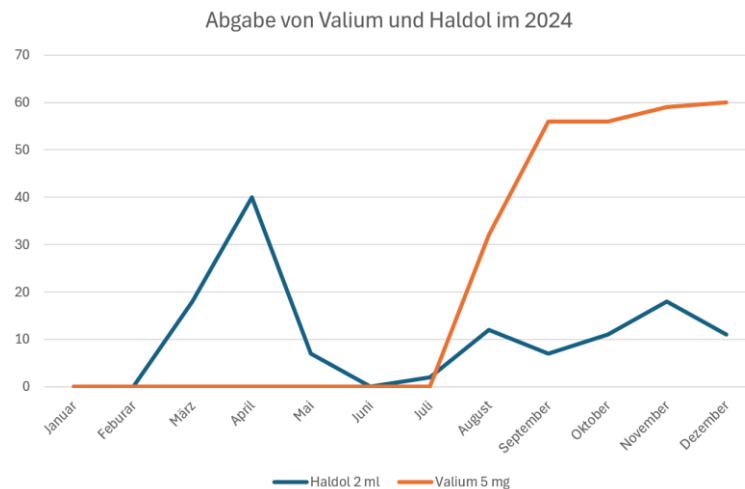


Abbildung 31: Abgabe von *Valium* und *Haldol* im Jahr 2024. Eigene Abbildung.

- Die Zunahme der Medikamentabgaben wurde dem LLM zurückgespielt. Entsprechend hat das Modell eine Auswertung gemacht. *ChatGPT* hat dabei Angaben wie «Anzahl Medikamente» und «Anzahl abgegebener Dosen» nicht korrekt wiedergegeben. Zudem wurden pflegerische Inhalte und Medikamente vermischt [63].

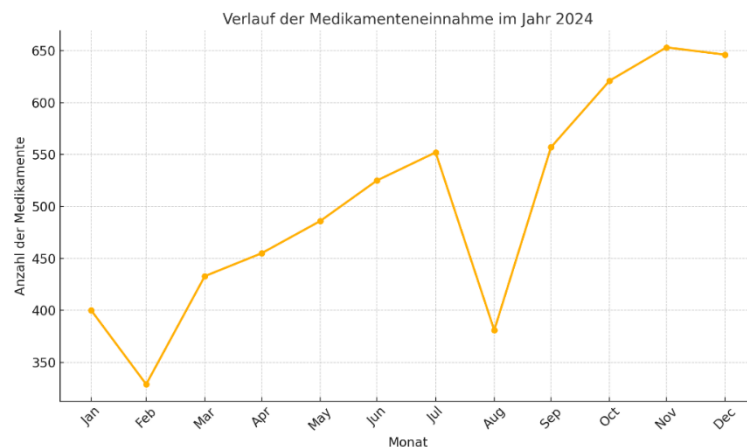


Abbildung 32: Fehlerhafter Verlauf der Medikamenteneinnahmen im Jahr 2024 [63].

- Aussagen wie «gelegentliche Übelkeit und verringertes Hungergefühl» [62, S. 9] oder «Nebenwirkungen wie Müdigkeit wurden beobachtet und im Team regelmässig evaluiert» [63, S. 4], wurden ohne valide Datenbasis gemacht. Auf Nachfrage erklärte die KI, dass «die Einschätzung «Gelegentliche Übelkeit und verringertes Hungergefühl» ... auf Grundlage allgemeiner Nebenwirkungen der in den Medikationstabellen genannten Medikamente getroffen [wurde]» [62, S. 9].
- In den Berichten des Dossiers *K-2021-00003* wurden alters- und krankheitsbedingte Veränderungen teilweise als positive Entwicklungen beschrieben, obwohl gegenteilige Informationen vorlagen.
- Einträge wie «Die regelmässige Beobachtung und Anpassung der unterstützenden Massnahmen trugen positiv zur Lebensqualität bei, da Lukas mehr Selbstvertrauen und Zufriedenheit in seine Alltagsgestaltung integrieren konnte» ist sehr hypothetisch und kann aufgrund der Akte nicht generalisiert werden [61, S. 4].

- Berichte enthielten keine detaillierten Informationen zu selbstverletzendem Verhalten oder dem gezielten Einsatz von Psychopharmaka in Krisensituationen. Bei K-2021-00079 wurde beispielsweise das selbstverletzende Verhalten erst auf Nachfrage erwähnt: *«Im Verlauf des Jahres wurden einzelne Episoden selbstverletzenden Verhaltens beobachtet, die in stressreichen Situationen auftraten. Durch den Einsatz von Deeskalationstechniken und gezielter Betreuung konnte in diesen Momenten schnell Stabilität hergestellt werden.»* [65, S. 4].

Hinweis: Das selbstverletzende Verhalten und die Pflege der entsprechenden Körperstelle werden im Journal häufig erwähnt. Die Aussage, dass die Stabilität durch den Einsatz von Deeskalationstechniken und gezielter Betreuung schnell wiederhergestellt werden konnte, lässt sich anhand der Journaldaten nicht validieren.

- Klinikaufenthalte oder Rückkehrsituationen erschienen nur in den Berichten, wenn sie in den Prompts ausdrücklich angefordert wurden [62].
- Das Modell konnte wichtige Bezugspersonen im Bericht nicht aufführen. Wenn *ChatGPT* danach gefragt wurde, erschienen zuerst alle Verfasserinnen bzw. Verfasser als Bezugspersonen. Nach Anpassung des Prompts wurden neben dem Namen des Klienten auch unpassende Antworten wie beispielsweise *am Mittag* oder *die Bauchfalte* geliefert, da diese Wörter oft in der Akte vorgekommen sind [66].
- In den Berichten wurden Begriffe und Formulierungen verwendet, die in der Sozialen Arbeit unüblich erscheinen. Statt neutraler und präziser Formulierungen fanden sich oft unprofessionelle Floskeln oder unpassende Aussagen:
 - **Beispiel 1:** *«Herr Keller hat die Wohngruppe routiniert besucht.»* Eine Fachperson würde *routiniert besucht* nicht schreiben. Herr Keller lebt auf der Wohngruppe [65, S. 3].
 - **Beispiel 2:** Floskeln wie *«Auffälligkeiten wurden im Team besprochen»* waren nicht nur inhaltsleer, sondern auch verwirrend, da sie keine konkrete Information lieferten [65, S. 4].

Weitere zentrale Schwachpunkte waren mangelnde Transparenz, technische Schwierigkeiten und fehlende Nachvollziehbarkeit der generierten Ergebnisse. Die Berichte enthielten Aussagen, deren Quelle in den hochgeladenen Dossiers nicht eindeutig nachvollziehbar war. Selbst bei gezielten Nachfragen konnte die KI ihre Entscheidungen nicht ausreichend erklären:

- Auf Nachfrage nach der Grundlage für *«David war an stressreichen Tagen weniger kooperativ»* lieferte die KI nur allgemeine Hinweise, ohne konkrete Quellen aus den Journaleinträgen zu benennen. Die Quellen konnten auch mit konkreten Rückfragen nicht eruiert werden [65, S. 13].
- Bei Nachfragen, um Aussagen zu überprüfen, generierte *ChatGPT* Python-Skripte. Diese durchsuchten die Journale lediglich nach Keywords wie *Stress* oder *Verweigerung*. Es zeigte sich, dass die KI selbst nicht in der Lage war, ihre Aussagen mit den hochgeladenen Daten konsistent und kontextbezogen abzugleichen.
- Die KI erzeugte gelegentlich Fehlermeldungen oder technische Ergebnisse, die für Fachpersonen der Sozialen Arbeit ohne technisches Hintergrundwissen nicht verständlich sind [60].
- Auf identische Prompts lieferte die KI oft unterschiedliche Antworten, was die Reproduzierbarkeit der Berichte stark einschränkte [65], [67].

Hinweis: Es ist bekannt, dass es API-Parameter für die Steuerung der Reproduzierbarkeit gibt, diese wurden in diesem Kontext allerdings nicht angewendet [68].

- Es ist unklar, in welchem Umfang die Datenqualität der Eingabedaten die Ergebnisse beeinflusst. Auch das Bereitstellen strukturierter Informationen über Medikamente konnte die Ergebnisqualität nicht signifikant verbessern [63], [69].

8.4.3 Bias-Analyse und Reduktion

In den generierten Berichten traten inhaltliche Verzerrungen auf, die sowohl die Gewichtung als auch die Darstellung bestimmter Informationen beeinflussten. Verzerrungen dieser Art stellen in der Sozialen Arbeit ein erhebliches Risiko dar, da sie die Realität verfälschen und zu fehlerhaften Schlussfolgerungen führen können. Beispiele dafür:

- In mehreren Berichten wurden Entwicklungen und Ereignisse positiver beschrieben als in den zugrunde liegenden Daten ersichtlich. Beispielsweise wurden altersbedingte oder krankheitsbedingte Verschlechterungen in Dossier *K-2021-00003* als *Fortschritte* interpretiert. Dadurch entstand der Eindruck einer positiven Entwicklung, obwohl in den Daten gegenteilige Tendenzen dokumentiert waren [62].
- Selbstverletzendes Verhalten und der gezielte Einsatz von Psychopharmaka wurden in den Berichten nur vereinzelt erwähnt und teilweise verkürzt dargestellt. Erst nach gezielten Rückfragen, etwa durch Prompts wie *«Berücksichtige selbstverletzendes Verhalten im Bericht»*, konnten diese Themen integriert werden. Diese Auslassungen führten zu unvollständigen und potenziell irreführenden Berichten [65].
- Klinikaufenthalte oder Anpassungen bei der Medikation wurden in einigen Berichten nicht erwähnt, während andere Details, die für die Fallanalyse weniger relevant sind, stärker betont wurden. Diese falsche Gewichtung erschwerte es, die Berichte als Grundlage für eine fundierte Fallanalyse zu verwenden [60].

8.4.4 Evaluierung der Ergebnisqualität

Die durch *ChatGPT* generierten Berichte konnten die Anforderungen an eine valide und professionelle Berichterstellung nicht erfüllen. Die Berichte enthielten Fehler, Auslassungen und nicht nachvollziehbare Informationen. Die Qualität wurde durch Halluzinationen, Beschönigungen und mangelnde Transparenz beeinträchtigt. Besonders problematisch war die Schwierigkeit, nachzuvollziehen, wie die KI zu ihren Aussagen kam, was die kritische Prüfung der Berichte erschwerte und zusätzlichen Aufwand erforderte.

Ein weiteres Risiko zeigte sich in der potenziellen Beeinflussung von Fachpersonen durch die generierten Inhalte. Ohne fundierte Dossier-Kompetenz besteht die Gefahr, dass fehlerhafte Schlussfolgerungen übernommen oder Halluzinationen unkritisch akzeptiert werden. Dies verdeutlicht die Notwendigkeit, dass Fachpersonen sowohl die Inhalte der Dossiers als auch die Funktionsweise der KI genau kennen müssen, um die Ergebnisse bewerten und korrigieren zu können.

Ein entscheidender Faktor für die Qualität der Berichte war das Prompting. Es zeigte sich, dass präzise und fachlich fundierte Eingaben notwendig sind, um relevante und korrekte Inhalte zu erhalten. Ohne entsprechende Vorgaben blieben die Ergebnisse oft ungenau, und selbst gezielte Prompts führten nur bedingt zu besseren Berichten.

Zusätzlich traten technische Herausforderungen auf, darunter schwer verständliche Fehlermeldungen und uneinheitliche Abläufe bei der Berichterstellung. Die KI generierte teils einfache Inline-Berichte, teils detaillierte Analysen, die durch unnötige oder verwirrende Zusatzinformationen auffielen. Diese Uneinheitlichkeit erschwerte die Nutzung und Nachvollziehbarkeit der Ergebnisse.

Auch die Abhängigkeit der Berichtqualität von den bereitgestellten Daten wurde deutlich. Unstrukturierte oder fehlerhafte Informationen wirkten sich negativ auf die Ergebnisse aus. Gleichzeitig gestaltete sich die Evaluierung als sehr zeitaufwendig, da die generierten Inhalte sorgfältig geprüft werden

mussten. Dies war insbesondere dann der Fall, wenn die KI Annahmen traf – etwa über allgemeine Nebenwirkungen von Medikamenten – obwohl diese nicht explizit in den Dossiers erwähnt waren.

Zusammenfassend bleibt festzustellen, dass die generierten Berichte für den professionellen Einsatz in der Sozialen Arbeit nicht geeignet sind. Die beschriebenen Schwächen – von Halluzinationen und inhaltlichen Lücken bis hin zu technischen Barrieren – überwiegen die potenziellen Vorteile. Um KI-Systeme in diesem Bereich effektiv nutzen zu können, sind erhebliche Verbesserungen erforderlich, sowohl in der technischen Entwicklung als auch in der Einbindung fachlicher Expertise.

8.5 Zusammenfassung und Fazit

Die Wahrnehmung von KI in der Praxis – sei es in Gesprächen mit Institutionen oder auf Fachveranstaltungen – ist häufig stark vereinfacht. Während einer Präsentation am *INSOS-Kongress 2023* wurde KI als eine Technologie vorgestellt, die mit minimalem Aufwand grosse Mehrwerte liefern kann: Daten werden erfasst, von einer KI verarbeitet und liefern unmittelbar Mehrwert [70]. Scherpenberg illustriert diese verbreitete, aber vereinfachte Wahrnehmung von KI in Unternehmen anhand einer schematischen Darstellung. Die folgende Abbildung zeigt diese vereinfachte Sichtweise, bevor in einer zweiten Grafik die tatsächlichen technologischen und organisatorischen Herausforderungen dargestellt werden [71]:

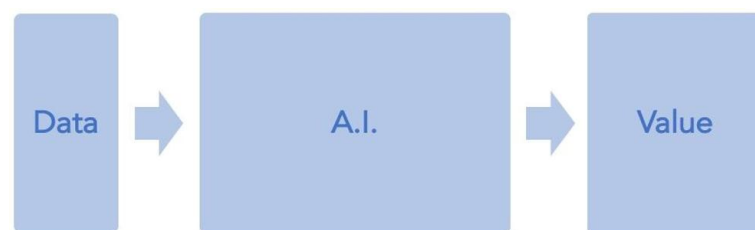


Abbildung 33: KI-Visualisierung «What companies think A.I. looks like» nach Scherpenberg [71].

Die Analyse in diesem Kapitel zeigt, dass diese vereinfachte Vorstellung der Realität nicht gerecht wird und der erfolgreiche Einsatz von KI weitaus komplexer ist. KI-gestützte Systeme erfordern klare Zielsetzungen, eine strukturierte Datenbasis, geeignete Modelle sowie präzise abgestimmte Prozesse, um eine nachhaltige Integration in bestehende Systeme zu ermöglichen. Die in diesem Kapitel untersuchten Herausforderungen – darunter Datenqualität, Spezifikationen, Transparenz und die Abhängigkeit von spezifischen Prompts – verdeutlichen, dass der Einsatz von KI keineswegs trivial ist. Die folgende Grafik von Scherpenberg stellt diese Komplexität dar, indem sie zusätzliche Faktoren einbezieht [71]:

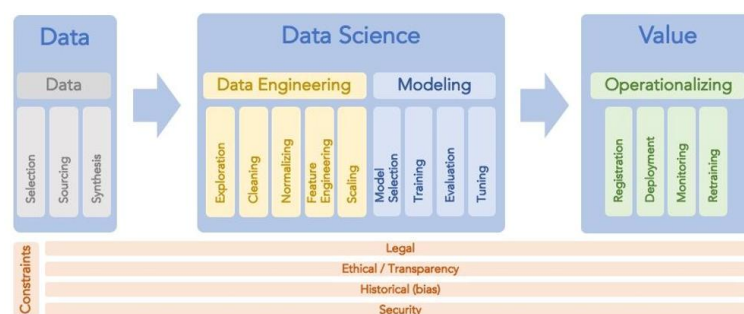


Abbildung 34: KI-Visualisierung «What it actually is» nach Scherpenberg [71].

Scherpenberg zeigt in der Grafik zusätzliche Abhängigkeiten zu rechtlichen Fragestellungen, Ethik und Transparenz, Verzerrungen und Sicherheit auf. Daneben sind auch fachliche Standards, Domänenwissen, Mitarbeitende und ihre Kompetenzen sowie Schulungen weitere wichtige Themen, welche

bei der Verwendung von KI adressiert werden müssen. Diese Faktoren beeinflussen die Planung, Entwicklung und Implementierung von KI-Lösungen und erfordern besondere Aufmerksamkeit, um nachhaltige und effektive Ergebnisse zu erzielen.

Ein ähnlicher Ansatz findet sich in der nachstehenden Grafik von Farah, die den MLOps-Ansatz (Machine Learning Operations) darstellt. Sie spiegelt die im Projekt gemachten Erfahrungen noch präziser wider und verdeutlicht, dass jede Phase – von der Datenverarbeitung über die Modellentwicklung bis hin zur Operationalisierung – ein iterativer Prozess ist. Jede einzelne Phase schafft dabei nicht nur einen Mehrwert für sich selbst, sondern auch für die gesamte Wertschöpfungskette. Durch diese Iterationen wird eine kontinuierliche Verbesserung ermöglicht, die sowohl technische als auch fachliche Anforderungen berücksichtigt [72]. Die von Scherpenberg benannten Abhängigkeiten (Constraints) sollten ergänzend ebenfalls auf dieses Modell angewendet werden, um den Erfolg langfristig sicherzustellen [71].

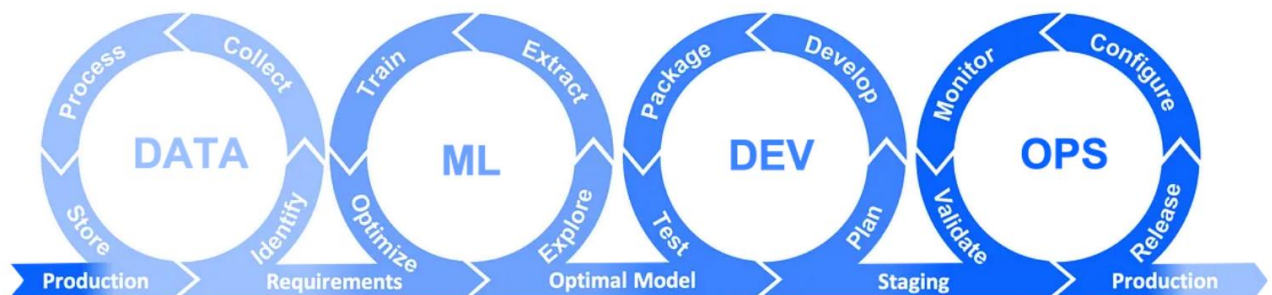


Abbildung 35: MLOps nach Farah [72].

Die Grafik verdeutlicht, dass die Integration von KI kein linearer Prozess ist. Vielmehr handelt es sich um einen von Beginn an festgelegten, sich wiederholenden und lernenden Kreislauf. Dieser Ansatz stellt höchste Ansprüche an die technische und fachliche Kompetenz aller Beteiligten. Nur durch enge Zusammenarbeit und kontinuierliche Anpassung der Methoden und Prozesse lässt sich langfristig ein Mehrwert erzielen.

Zusammenfassend zeigt sich, dass der fundierte Einsatz von KI weit über vereinfachte Vorstellungen hinausgeht, wie sie oft in der Praxis vertreten werden. Er erfordert klare Zielsetzungen, robuste Datenstrukturen, iterative Prozesse und eine enge Zusammenarbeit zwischen Fachkräften und Technik. Ohne diese Voraussetzungen bleibt das Potenzial von KI begrenzt und schwer realisierbar.

8.5.1 Erkenntnisse zur Zielsetzung

Die Ergebnisse zeigen, dass eine KI-gestützte Berichterstellung grundsätzlich möglich ist, jedoch mit erheblichen Herausforderungen verbunden bleibt. Während *ChatGPT* mit *GPT-4o* eine hohe sprachliche Qualität in der Textgenerierung aufweist und durch gezieltes Prompting überzeugende Berichte erstellen kann, traten in der Evaluation signifikante Schwächen in der Reproduzierbarkeit, Transparenz und fachlichen Korrektheit auf.

Ein zentrales Problem war die Neigung des Modells zu Halluzinationen. Die KI fügte nicht vorhandene Informationen in die Berichte ein oder interpretierte Sachverhalte fehlerhaft. Zudem war eine Datenanonymisierung zwingend erforderlich, um die Datenschutzanforderungen zu erfüllen. Dies führte jedoch dazu, dass wesentliche Kontexte aus den Dokumenten entfernt wurden, wodurch die Qualität der generierten Berichte weiter beeinträchtigt wurde. Die unstrukturierte Datenbasis stand in Zusammenhang mit variierenden Ergebnissen. Eine Wiederholung der Untersuchungen mit einer strukturierten Datenbasis könnte zusätzliche Erkenntnisse liefern.

Die Resultate deuten darauf hin, dass KI als unterstützendes Werkzeug genutzt werden kann, jedoch ohne Nachbereitung durch Fachpersonen unzuverlässig bleibt. Damit stehen die Erkenntnisse im Einklang mit aktuellen Forschungsperspektiven, die betonen, dass KI-basierte Systeme eine fachliche Begleitung und institutionelle Standards benötigen, um in sensiblen Bereichen wie der Sozialen Arbeit adäquat eingesetzt werden zu können.

9 Wirtschaftliche Ergebnisse

Die technischen Ergebnisse zeigen, dass die Qualität der zugrunde liegenden Daten einen entscheidenden Einfluss auf die Leistungsfähigkeit einer KI-gestützten Berichterstellung hat. Ohne eine strukturierte und konsistente Datenbasis sind weder eine zuverlässige technische Umsetzung noch belastbare wirtschaftliche Analysen möglich. Ursprünglich war vorgesehen, eine wirtschaftliche Perspektive zu erarbeiten, welche die Kosten für Entwicklung, Betrieb und Wartung berücksichtigt. Aufgrund der unzureichenden Datenqualität ist eine belastbare Quantifizierung dieser Faktoren jedoch nicht möglich. Stattdessen wird im Folgekapitel untersucht, welche Ansätze zur Verbesserung der Datenqualität bestehen.

Mehrere Unsicherheiten verhindern derzeit eine wirtschaftliche Bewertung. Die Kosten für eine Optimierung der Datenqualität lassen sich nicht verlässlich abschätzen, da die Verantwortung hierfür bei den Institutionen liegt und keine einheitlichen Massnahmen definiert sind. Zudem fehlen belastbare Vergleichswerte, da bislang keine funktionierende KI-Integration mit realen Testszenarien vorliegt. Auch die Betriebs- und Wartungskosten sind erst dann valide bewertbar, wenn feststeht, ob die KI lokal oder cloudbasiert implementiert wird. Solange diese grundlegenden Fragen offen sind, bleibt eine wirtschaftliche Bewertung nicht möglich. Erst mit einer verlässlichen Datenbasis und klar definierten technischen Rahmenbedingungen kann eine fundierte ökonomische Analyse erfolgen.

10 Diskussion

Die fehlende Datenqualität und die daraus resultierenden Unsicherheiten verdeutlichen, dass der erfolgreiche Einsatz von KI-gestützter Berichterstellung nicht nur eine technologische, sondern auch eine organisatorische Herausforderung darstellt. Um eine nachhaltige Implementierung zu ermöglichen, müssen bestehende Probleme gezielt adressiert und Optimierungsmöglichkeiten erarbeitet werden. In diesem Kapitel werden daher die zentralen Herausforderungen sowie konkrete Massnahmen zur Verbesserung der Datenqualität, Modellleistung und Integration von KI in bestehende Systeme diskutiert.

10.1 Einfluss der Datenqualität auf MLOps

Die Forschung belegt, dass die Leistung von KI-Modellen massgeblich von der Qualität der zugrunde liegenden Daten abhängt (siehe Kapitel 4.2.2 *Trainingsdaten und ihre Implikationen* und 4.2.4 *Qualität der Eingabedaten*). Gleichzeitig haben die Analysen in Kapitel 8.3 *Datenanalyse* und 8.4 *Erstellung von Berichten* gezeigt, dass die aktuellen Daten erhebliche Defizite in Bezug auf Konsistenz, Strukturierung und Nachvollziehbarkeit aufweisen. Dies legt nahe, dass die mangelhafte Datenqualität einen wesentlichen Einfluss auf die unzureichenden Ergebnisse der KI-gestützten Berichterstellung hat. Um die Modellleistung wissenschaftlich fundiert bewerten zu können, muss die Qualität der Daten systematisch verbessert und entlang des MLOps-Prozesses gezielt optimiert werden. Erst nach einer solchen Optimierung lassen sich weiterführende Tests durchführen, um belastbare Aussagen über die tatsächliche Eignung von KI-Modellen für die Berichterstellung in der Sozialen Arbeit zu treffen. Zur Verdeutlichung wird nachfolgend erneut die Grafik von Farah dargestellt (siehe auch Kapitel 8.5 *Zusammenfassung und Fazit*). In diesem Abschnitt wird sie jedoch unter dem Gesichtspunkt der technischen Optimierung entlang des MLOps-Prozesses analysiert.

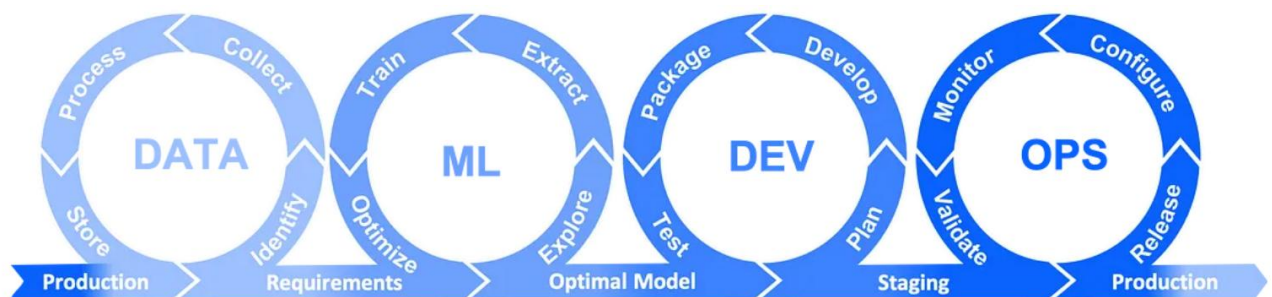


Abbildung 36: MLOps nach Farah [72].

Der folgende Abschnitt identifiziert zentrale Herausforderungen entlang der MLOps-Phasen und unterscheidet zwischen übergeordneten Rahmenbedingungen sowie zyklischen Prozessschritten. Da die Qualität der Daten die Leistungsfähigkeit eines KI-Modells direkt beeinflusst, stehen die datenbezogenen Herausforderungen im Zentrum der ersten Phasen. Die folgende Tabelle zeigt, welche Probleme entlang des Datenflusses (Data) auftreten und welche strukturellen Defizite derzeit bestehen.

Herausforderungen	
Identify	- Die Analyse basiert nur auf Daten einer Institution, wodurch die Repräsentativität stark eingeschränkt ist. - Es ist unklar, welche weiteren relevanten Datenquellen (z. B. Falldokumentationen, externe Berichte) existieren und genutzt werden könnten.

Collect	• Uneinheitliche Erfassung: Fachpersonen dokumentieren mit unterschiedlichen Detailgraden und Schreibweisen, was die Konsistenz beeinträchtigt. • Fehlende Prozesse zur Qualitätssicherung: Ungenaue oder unvollständige Einträge werden nicht überprüft. • Es braucht eine Klärung, in welchen Modulen Inhalte gespeichert werden.
Process	• Uneinheitliche Kategorisierung: Manche Einträge sind nur grob kategorisiert (<i>Beobachtung</i>), während spezifische Themen wie <i>Pflege</i> oder <i>Medikation</i> nicht konsequent verwendet werden. • Fehlende Standardisierung von Markierungen erschwert die Filterung und gezielte Analyse von Daten. • Es gibt keine automatisierte Datenbereinigung, um Tippfehler oder inkonsistente Begriffe zu korrigieren. • Fehlende Standardisierung von Berichten. • Uneinheitliche Datenanonymisierung der Einträge.
Store	• Die Speicherung der Daten erfolgt ausschliesslich innerhalb der Applikation <i>socialweb</i>, wodurch externe Analysen erschwert werden. • Es gibt keine Infrastruktur zur systematischen Extraktion und Verarbeitung der Daten in einem zentralen Data Warehouse, das Daten aus verschiedenen Quellen sammelt und für Analysen aufbereitet. • Der Zugriff auf sensible Daten für Validierungszwecke und Modelltraining ist nicht ausreichend geklärt.

Tabelle 17: Herausforderungen im MLOps-Prozess nach Farah im Bereich *Data*.

Während Datenqualität eine Grundvoraussetzung für erfolgreiche KI-Modelle ist, ergeben sich in den nachfolgenden Phasen der Modellentwicklung zusätzliche Herausforderungen. Insbesondere fehlen klare Anforderungen für die Bewertung der Modellleistung sowie objektive Vergleichskriterien. Auch grundlegende Rahmenbedingungen wie interdisziplinäre Abstimmung und akzeptable Fehlergrenzen sind noch unzureichend definiert. Besonders in sensiblen Anwendungsbereichen ist unklar, welche Fehlerquote vertretbar ist und welche Abweichungen akzeptiert werden können. Diese Unsicherheiten erschweren die gezielte Entwicklung, das Training und die Evaluation von Modellen. Die folgende Tabelle fasst die zentralen Herausforderungen im Bereich der Modellentwicklung (ML) zusammen:

Herausforderungen	
Explore	• Die Evaluierung alternativer KI-Modelle ist unzureichend: Nur <i>GPT-4o</i> wurde getestet, weitere Modelloptionen fehlen. • Es existiert kein standardisiertes Referenz-Datenset oder spezifische Fachvorgaben zur objektiven Bewertung der Modellleistung. • Es muss geprüft werden, ob spezialisierte KI-Modelle für den Bereich der Sozialen Arbeit existieren – insbesondere für die deutsche, französische und italienische Sprache.
Extract	• Es fehlen einheitliche Standards zur strukturierten Datenerfassung in der Sozialen Arbeit, was die Merkmalsextraktion erschwert. • Die unstrukturierten Journaldaten führten zu einer hohen Varianz in der Kategorisierung und Terminologie. • Data Scientists und Fachkräfte arbeiten nicht eng genug zusammen, um Merkmale zu identifizieren.
Train	• Da kein optimiertes Modell für die Berichterstellung gefunden wurde, konnte kein gezieltes Training durchgeführt werden. • Die vorhandenen Trainingsdaten sind unstrukturiert und inkonsistent, wodurch eine effektive Anpassung der Modelle erschwert wird. • Es fehlen standardisierte Metriken zur Bewertung des Trainingserfolgs.
Optimize	• Fehlende Standards verhindern eine gezielte Identifikation von Verbesserungspotenzialen. • Ohne klar definierte Anforderungen bleibt eine systematische Optimierung unrealistisch.

Tabelle 18: Herausforderungen im MLOps-Prozess nach Farah im Bereich *ML*.

Die bisherigen Erkenntnisse zeigen, dass eine fundierte Modellbewertung derzeit nicht möglich ist. Weder eine belastbare Datenbasis noch definierte Modellanforderungen liegen vor, sodass eine Entscheidung über das optimale Modell aktuell nicht getroffen werden kann. Auch die Frage der Operationalisierung bleibt offen: Ohne eine verbesserte Datenqualität und grundlegende Tests sind weitere Optimierungen rein spekulativ. Unter diesen Bedingungen ist eine detaillierte Diskussion zur finalen Modellwahl und Implementierung nicht zielführend. Daher wird an dieser Stelle auf eine weitergehende Analyse der Entwicklungs- (DEV) und Betriebsphase (OPS) verzichtet.

10.2 Akteurinnen und Akteure im MLOps-Prozess

Die Analyse zeigt, dass viele Herausforderungen entweder unmittelbar durch mangelhafte Datenqualität verursacht werden oder daraus resultieren. Bevor weitere technische oder methodische Anpassungen sinnvoll evaluiert werden können, muss daher zwingend eine Optimierung der Datenbasis angestrebt werden. Die Verbesserung der Datenqualität erfordert eine enge Zusammenarbeit verschiedener Beteiligter. Die nachfolgende Grafik verdeutlicht, dass nicht nur die Institutionen und Softwareunternehmen eine Rolle spielen, sondern auch Externe wie Fachverbände, Forschungsinstitutionen und regulatorische Stellen. Ebenso sind die Klientinnen und Klienten indirekt betroffen, da ihre Bedürfnisse und Lebensrealitäten durch die Datenrepräsentation beeinflusst werden. Auch wenn sie nicht als eigenständige Akteursgruppe behandelt werden, müssen alle Massnahmen ihr Wohl und ihre Integrität sichern. Institutionen, Softwareunternehmen und Externe sollten stets im Interesse der betreuten Personen handeln und die ethischen sowie fachlichen Grundsätze der Sozialen Arbeit berücksichtigen. Entsprechend werden sie in der Grafik als eigenständige, jedoch nicht aktiv steuernde Gruppe in Grau dargestellt.

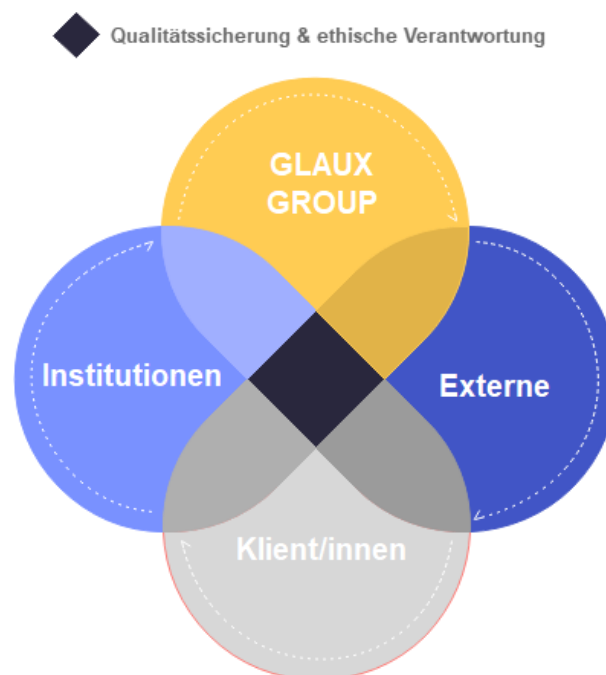


Abbildung 37: Beteiligte für die Optimierung der Datenbasis. Eigene Darstellung.

Jede dieser Gruppen hat eigene Verantwortlichkeiten und Einflussmöglichkeiten auf die Datenqualität. Während Institutionen für eine konsistente Erfassung und Strukturierung der Daten zuständig sind, kann die *GLAUX GROUP* durch technische Lösungen die Datenverarbeitung und -validierung unterstützen. Externe beeinflussen wiederum Standards, Forschung und regulatorische Rahmenbedingungen.

Die nächsten Kapitel legen den Fokus auf die Optimierung der Datenbasis als Grundlage für eine zukünftige Implementierung. Im Zentrum stehen dabei folgende Schwerpunkte:

- Technische Unterstützung der *GLAUX GROUP*
- Fachbezogene Lösungsansätze der Institutionen
- Einbindung externer Akteurinnen und Akteure

Diese Analyse bildet die Grundlage für die weitere Diskussion über notwendige Massnahmen zur Verbesserung der technischen Rahmenbedingungen und zeigt auf, welche Schritte erforderlich sind, um eine KI-gestützte Berichterstellung in der Sozialen Arbeit realistisch nutzbar zu machen.

10.2.1 GLAUX GROUP

Die bestehenden Softwarefunktionen bieten bereits gute Möglichkeiten, um die Datenqualität zu verbessern und die Daten gezielt für die KI-gestützte Berichterstellung aufzubereiten. Dieses Kapitel beschreibt die vorhandenen technischen Funktionen zur Datenstrukturierung und zeigt, wie sie gezielt für eine bessere Datenqualität genutzt werden können. Die nachfolgende Grafik zeigt die aktuelle Erfassungsmaske im *Journal*, welches hauptsächlich für Dokumentationszwecke in der täglichen Arbeit verwendet wird:

Journaleintrag erstellen

Journaleintrag für: Cadegg Seraina

Klassifizierung: 01 Vertraulich: Wohnen

Kategorie:

Markierungen:

Eintrag vom: 05.02.2025 09:23

In Kontakt mit: Kontakt suchen...

Inhalt

Rich text editor toolbar: Bold, Italic, Strikethrough, Bulleted list, Numbered list, Indent, Outdent, Table, Link, Unlink, Image, Video, Code, Undo, Redo, Textbausteine, Platzhalter.

Buttons: Abbrechen, Speichern & Neu, Speichern

Abbildung 38: Printscreen *Journaleintrag erstellen* in *socialweb*.

Die Funktionen, welche bereits heute eine technische Basis für die Verbesserung der Datenqualität darstellen, werden in den untenstehenden Tabellen aufgelistet und kurz beschrieben:

Feld / Funktion	Kardinalität	Beschreibung
Klassifizierungen	0:1	Können genutzt werden, um sensitive Inhalte gezielt zu kennzeichnen und den Zugriff darauf einzuschränken. Es kann keine oder eine Klassifizierung verwendet werden.
Kategorie	1:1	Erfordert eine eindeutige thematische Kategorisierung der Einträge – jeder Eintrag muss einer Kategorie zugeordnet werden.
Markierungen	0:n	Erlaubt eine mehrdimensionale Verknüpfung von mehreren Themen. Diese können sich inhaltlich unterscheiden. Beispiele dafür sind <i>Infos für Nachtwache</i> oder <i>Wichtig für Bericht</i> . Ein Eintrag kann mehreren relevanten Inhalten zugeordnet werden.
In Kontakt mit	0:n	Ermöglicht die Verknüpfung von Einträgen mit spezifischen Personen oder Institutionen, anstatt diese unstrukturiert im Freitext zu nennen. Ein Eintrag kann mehreren Personen oder Institutionen zugewiesen werden, um die Nachvollziehbarkeit zu verbessern.
Textbausteine		Dienen als vorgefertigte Formulierungen, um eine einheitliche und strukturierte Dokumentation zu erleichtern.
Rechtschreibprüfung		Die browserbasierte Rechtschreibprüfung verbessert die sprachliche Qualität.

Tabelle 19: Aktuelle Felder und Funktionen beim Erstellen von Journaleinträgen in *socialweb*.

Obwohl diese Funktionen vorhanden sind, werden sie bisher nur eingeschränkt zur Sicherstellung einer konsistenten Datenqualität eingesetzt. Eine strukturierte Nutzung dieser Optionen könnte eine wichtige Grundlage für eine spätere KI-gestützte Berichterstellung schaffen.

Der Autor kennt die Herausforderungen der Dokumentation aus der Praxis und ist sich bewusst, dass die verfügbare Zeit für die Erfassung und Validierung von Einträgen begrenzt ist. Eine digitale Unterstützung könnte helfen, den Dokumentationsaufwand zu reduzieren und gleichzeitig die Datenqualität zu verbessern. Um das Potenzial eines solchen Ansatzes gezielt zu evaluieren, wäre es sinnvoll, technische Funktionen zu prüfen, die Fachpersonen direkt bei der Datenerfassung entlasten. Die Idee ist, bereits bei der Eingabe strukturelle Vorgaben zu überprüfen. Entsprechend könnte KI bereits in diesem frühen Stadium eingesetzt werden, um folgende Aspekte beim Speichern des Journaleintrags zu prüfen:

Thema	Beschreibung
Kategorisierung	Automatische oder unterstützte Zuordnung zu bestehenden Kategorien.
Markierungen	Automatische Vorschläge für Markierungen basierend auf dem Inhalt.
Schreibstil	Erkennung langer oder unklarer Formulierungen und Vorschläge zur besseren Strukturierung.
Fachliche Validierung	Prüfung, ob der Eintrag mit den thematischen und inhaltlichen Vorgaben der Institution übereinstimmt.
Sprachliche Qualitätssicherung	Sicherstellung korrekter Grammatik und Rechtschreibung.

Inhaltliche Validierung	Sicherstellung, dass Informationen in den dafür vorgesehenen Modulen erfasst und eingetragen wurden. Prüfung auf mögliche Verzerrungen (Bias) und entsprechende Anzeige. Überprüfung, ob eine Aufteilung in mehrere Einträge sinnvoll ist.
Datenanonymisierung	Identifikation personenbezogener Daten und Vorschläge zur Anonymisierung oder Verknüpfung unter « <i>In Kontakt mit</i> » anstatt direkter Namensnennung.

Tabelle 20: Validierungsoptionen beim Speichern von Journaleinträgen in *socialweb*.

Durch eine gezielte Optimierung der Dateneingabe und -strukturierung kann die Datenqualität signifikant verbessert werden. Zusätzlich ermöglicht diese Funktionalität eine effizientere Datenverarbeitung. Anstatt vollständige Dokumentationen an die KI zu übermitteln, können einzelne, vorstrukturierte Einträge verarbeitet werden. Dies verbessert die Performance und reduziert die Verarbeitungszeit. Diese Anpassungen schaffen eine fundierte Basis für eine spätere automatisierte Berichterstellung und ermöglicht eine belastbare Wiederholung der Tests. Die wichtigsten Verbesserungen:

- **Höhere Validität der Daten**
Die strukturierte Erfassung reduziert Inkonsistenzen und verbessert die Nachvollziehbarkeit.
- **Automatisierte Anonymisierung**
Eine frühzeitige Identifikation und Anonymisierung personenbezogener Daten erleichtert die Modellauswahl und minimiert datenschutzrechtliche Hürden.
- **Erstinstanzliche Bias-Prüfung**
Durch strukturierte Vorgaben kann bereits frühzeitig eine kritische Prüfung von Verzerrungen erfolgen.
- **Menschliche Validierung**
Bereits bei der Datenerfassung kann eine erste menschliche Prüfung erfolgen, wodurch potenzielle Fehler oder Verzerrungen direkt erkannt und korrigiert werden können.

Die hier aufgezeigten Optimierungsmöglichkeiten ergänzen die bestehenden technischen Möglichkeiten, die bereits eine solide Grundlage für eine bessere Datenqualität bieten. Damit dieses Potenzial jedoch vollständig ausgeschöpft werden kann, müssen die Institutionen zunächst strukturelle Anpassungen vornehmen. Die Einführung klarer Dokumentationsrichtlinien und die Schulung der Mitarbeitenden sind essenziell, um eine einheitliche und qualitativ hochwertige Datenerfassung sicherzustellen. Erst wenn diese Grundlagen geschaffen sind, können die hier aufgezeigten technischen Verbesserungen dazu beitragen, die Prozesse weiter zu professionalisieren und zu optimieren.

10.2.2 Institutionen

Die Einführung KI-gestützter Berichterstellung setzt voraus, dass die erfassten Daten strukturiert, nachvollziehbar und konsistent sind. Dies kann jedoch nur gelingen, wenn die Institutionen selbst klare Vorgaben und Prozesse für die Datenerfassung definieren und implementieren. Wie bereits in Kapitel 8.3.3 *Zusammenfassung und Handlungsempfehlungen* diskutiert, sind standardisierte Dokumentationsprozesse entlang fachlicher Konzepte für eine verlässliche Datenbasis entscheidend. Die folgenden Massnahmen zielen darauf ab, diese strukturellen Defizite zu beheben und eine konsistente, für KI-Anwendungen geeignete Datenerfassung sicherzustellen. Institutionen sollten daher klare Leitlinien entwickeln, die definieren, welche Informationen für Berichte benötigt werden und wie sie zu dokumentieren sind. Dabei könnte ein Rahmenmodell helfen, das sich an den typischen Anforderungen von Berichten orientiert. Dies würde sicherstellen, dass alle relevanten Aspekte systematisch erfasst werden, ohne dass nachträgliche Korrekturen oder Ergänzungen erforderlich sind. Bei der konzeptionellen Ausarbeitung neuer Dokumentationsrichtlinien, müssen auch Fachpersonen einbezogen werden. Technische Optimierungen können nur dann wirksam werden, wenn die Anforderungen verstanden und in die Arbeitsabläufe integriert sind. Die Einführung neuer Dokumentationsstrukturen darf daher nicht als reine Top-down-Anweisung verstanden werden. Eine nachhaltige Veränderung

gelingt nur, wenn die Fachpersonen selbst die Notwendigkeit erkennen und mittragen. Eine erfolgreiche Umsetzung erfordert daher gezielte Sensibilisierung und Schulung:

- **Information**
Vermittlung der Notwendigkeit einer strukturierten Datenerfassung und der Vorteile für die eigene Arbeit.
- **Schulung**
Praktische Trainings zur Nutzung technischer Hilfsmittel wie Kategorisierung, Markierungen und Anonymisierung.
- **Begleitung**
Langfristige Unterstützung durch Multiplikatorinnen und Multiplikatoren oder interne Fachpersonen zur Qualitätssicherung.
- **Feedbackschleifen**
Regelmässige Feedbackschleifen, beispielsweise durch Workshops oder Umfragen, ermöglichen eine frühzeitige Identifikation von Herausforderungen und gezielte Anpassungen.

Besonders herausfordernd ist dabei, dass nicht nur die operativen Mitarbeitenden, sondern auch die Leitungspersonen oft nur über begrenzte IT-Kenntnisse verfügen. Eine erfolgreiche Implementierung erfordert daher eine durchgängige Schulungsstrategie auf allen Ebenen der Institution. Ein einmalig definiertes Konzept wird zudem nicht ausreichen, da sich Anforderungen an Dokumentation und Berichterstellung weiterentwickeln. Insbesondere auch angestossen durch den Lernprozess der Institution selbst, soll ein iterativer Prozess entstehen, in dem regelmässig überprüft wird, ob die strukturierten Erfassungsmethoden praktikabel und ob Anpassungen erforderlich sind.

Neben der strukturierten Datenerfassung müssen Institutionen auch sicherstellen, dass ihre Mitarbeitenden über ausreichende IT- und KI-Kompetenzen verfügen. Eine grundlegende IT-Affinität ist erforderlich, um technische Hilfsmittel effektiv zu nutzen. Darüber hinaus erfordert die Einführung KI-gestützter Berichterstellung ein Verständnis für die Funktionsweise und Grenzen solcher Systeme. Dazu gehören insbesondere:

- **IT-Grundkompetenzen**
Sicherer Umgang mit Hard- und Software, Dokumentationssystemen und Datenverarbeitung.
- **Datenbewusstsein**
Verständnis für Datenqualität, Datenschutz und Anonymisierung.
- **KI-Grundlagen**
Sensibilisierung für Chancen und Risiken KI-gestützter Prozesse, inkl. einer ethisch-kritischen Reflexion der Ergebnisse.

Die technische Unterstützung kann nur dann wirksam werden, wenn die Institutionen parallel dazu klare Strukturen für die Datenerfassung schaffen und die Mitarbeitenden aktiv in den Prozess einbinden. Dies erfordert nicht nur neue Konzepte und Schulungen, sondern auch eine langfristige Haltung, die die Bedeutung qualitativ hochwertiger Daten anerkennt. Eine erfolgreiche KI-gestützte Berichterstellung setzt eine enge Verzahnung technischer Lösungen mit klaren fachlichen Standards voraus. Nur wenn beide Bereiche ineinandergreifen, können valide, nachvollziehbare und praxisnahe Ergebnisse erzielt werden.

10.2.3 Externe Akteurinnen und Akteure

Die Optimierung der Datenqualität und die Einführung KI-gestützter Berichterstellung sind nicht allein Aufgaben der Institutionen oder Softwareunternehmen. Eine nachhaltige und fachlich fundierte Umsetzung erfordert die Einbindung externer Akteurinnen und Akteure, welche mit ihren jeweiligen

Kompetenzen und Zuständigkeiten einen wichtigen Beitrag leisten. Anbei eine Auflistung der Beteiligten, welche die Qualität der Ergebnisse weiterhin stark beeinflussen könnten:

- **Hochschulen und Berufsbildung**
Die Vermittlung von IT- und KI-Grundlagen sollte einen festen Platz in der Ausbildung und Weiterbildung von Fachkräften der Sozialen Arbeit haben. Höhere Fachschulen und Fachhochschulen sollten sicherstellen, dass Studierende grundlegende Kompetenzen im Bereich strukturierte Datenerfassung, Datenschutz und digitale Werkzeuge erwerben. Gleichzeitig könnten sie Weiterbildungsangebote für bereits tätige Fachpersonen schaffen, die niederschwellige Möglichkeiten bieten, sich fundierte Kenntnisse in diesen Bereichen anzueignen respektive weiter zu vertiefen. Neben der Ausbildung haben Hochschulen auch die Aufgabe, den wissenschaftlichen Diskurs über fachliche Standards zu führen. Sie sollten gemeinsam mit kantonalen Stellen, politischen Entscheidungsträgern und Fachorganisationen entsprechende Leitlinien entwickeln und fördern.
- **Fachverbände und Qualitätssicherungsstellen**
Berufsverbände der Sozialen Arbeit haben eine zentrale Rolle bei der Entwicklung und Durchsetzung professioneller Standards. Sie können Empfehlungen für den Einsatz KI-gestützter Berichterstellung formulieren und sicherstellen, dass neue Technologien mit den fachlichen und ethischen Grundsätzen der Sozialen Arbeit in Einklang stehen. Zudem könnten Zertifizierungs- und Qualitätssicherungsstellen offizielle Rahmenbedingungen für den Einsatz solcher Technologien schaffen. Dies würde dazu beitragen, dass sich Mindeststandards für Datenqualität und Dokumentation branchenweit durchsetzen.
- **Datenschutz- und Aufsichtsbehörden**
Da KI-gestützte Systeme häufig personenbezogene und sensible Daten verarbeiten, sollten Datenschutzbehörden frühzeitig in den Entwicklungsprozess einbezogen werden. Sie können helfen, kritische Aspekte wie Datenanonymisierung, Zugriffsrechte und Speicherfristen zu bewerten und sicherzustellen, dass regulatorische Vorgaben eingehalten werden. Ein enger Austausch mit diesen Stellen könnte zudem sicherstellen, dass neue technische Lösungen datenschutzkonform gestaltet sind und potenzielle Risiken frühzeitig erkannt werden.
- **Softwareunternehmen und interdisziplinäre Zusammenarbeit**
Eine branchenweite Zusammenarbeit zwischen verschiedenen IT-Unternehmen könnte dazu beitragen, blinde Flecken zu reduzieren und gemeinsame Standards für die Verarbeitung von Daten in der Sozialen Arbeit zu entwickeln. Darüber hinaus wäre eine verstärkte interdisziplinäre Zusammenarbeit mit Forschungseinrichtungen ausserhalb der Hochschulen, etwa unabhängigen Instituten für KI, Ethik oder Sozialwissenschaften, sinnvoll. Sie könnten neue Erkenntnisse in die Entwicklung einbringen und dazu beitragen, praxisnahe und sozialverträgliche Lösungen zu erarbeiten.
- **Finanzierungsstellen und öffentliche Hand**
Auch Institutionen wie Stiftungen, Krankenkassen oder staatliche Finanzierungsstellen haben ein Interesse an standardisierten und nachvollziehbaren Berichtsprozessen. Sie könnten gezielt Projekte fördern, die zur Qualitätssicherung und Weiterentwicklung von Dokumentationsstandards beitragen. Indem sie Qualitätskriterien für die Finanzierung von Institutionen an standardisierte Berichterstattung knüpfen, könnten sie zudem einen Anreiz für eine breitere Einführung strukturierter Datenprozesse schaffen.

Die Verbesserung der Datenqualität und die Einführung KI-gestützter Berichterstellung sind komplexe Prozesse, die über die einzelne Institution hinausgehen. Eine enge Zusammenarbeit mit externen Akteurinnen und Akteuren kann dazu beitragen, fachliche Standards zu etablieren, regulatorische Sicherheit zu gewährleisten und innovative technologische Entwicklungen gezielt zu fördern. Nur durch eine koordinierte, interdisziplinäre Herangehensweise lassen sich nachhaltige und praxistaugliche Lösungen realisieren.

10.2.4 Zusammenfassung und Fazit

Die vorherigen Abschnitte haben gezeigt, dass die Einführung von KI-gestützter Berichterstellung eine enge Zusammenarbeit zwischen verschiedenen Akteurinnen und Akteuren erfordert. Besonders die Verzahnung von technischer Entwicklung, ethischen Standards und praxisnahen Evaluierungen stellt eine zentrale Herausforderung dar. Die folgende Tabelle fasst die zentralen Aufgaben entlang des MLOps-Prozesses zusammen und zeigt mögliche Perspektiven für zukünftige – idealerweise gemeinsame – Entwicklungen und Standards auf.

Phase	Schritt	GLAUX GROUP	Institutionen	Externe
DATA	Identify	Definition der relevanten Datenquellen in socialweb.	Identifikation und Bereinigung relevanter Daten.	Erarbeitung von Fachstandards für die Erhebung und Speicherung relevanter Daten.
	Collect	Zusammenführung relevanter Daten aus bestehenden Modulen.	Sammlung und Kategorisierung von Daten nach definierten Standards.	Überprüfung regulatorischer und ethischer Anforderungen für die Datennutzung.
	Process	(Automatisierte) Vorverarbeitung für ML-Training.	Sicherstellung der Qualität und Konsistenz der Datenerfassung.	Überprüfung der Einhaltung von Fachstandards.
	Store	Prüfung bestehender Datenbankstrukturen für die Nutzung von KI.		Definition von Standards für die Speicherung und Archivierung.
ML	Explore	Analyse der Datenqualität und Identifikation von Bias-Problemen.	Feedback zur Datenverarbeitung für den Praxiseinsatz.	Begleitung durch Forschungseinrichtungen zur Validierung.
	Extract	Feature Engineering für Modelle, Erstellung eines Golden Data Sets.	Definition der relevanten Datenpunkte für KI-gestützte Funktionen.	Empfehlung zu fachgerechten und nachhaltigen Datenverarbeitungsmethoden.
	Train	Training der ersten KI-Modelle auf bereinigten Daten, ergänzt durch synthetische Daten.	Pilotversuche mit Fachpersonen zur Qualitätssicherung.	Begleitung von Pilotversuchen zur Qualitätssicherung.
	Optimize	Optimierung und Fine-Tuning der Modelle.	Anpassung der Prozesse für eine bessere KI-Integration.	Festlegung von Standards für Transparenz und Nachvollziehbarkeit.
DEV	Plan	Technische Konzeption für Umsetzung.	Definition von Anforderungen für KI-gestützte Funktionen.	
	Develop	Programmierung der KI-Funktionen.		
	Package	Bereitstellung der KI-Modelle im bestehenden Ökosystem.		
	Test	Durchführung von automatisierten Unit- und Integrationstests.	Praktische Tests mit Fachpersonen.	Begleitung durch Forschungseinrichtungen zur Validierung.
OPS	Release	Rollout der ersten KI-gestützten Funktionen.	Einführung in Pilotinstitutionen.	
	Configure	Anpassung der KI an verschiedene Anwendungsfälle.	Schulungen für Fachkräfte zur Nutzung der KI.	

	Monitor	Laufende Qualitätskontrolle und Fehleranalyse.	Kontinuierliches Feedback aus der Praxis.	Regelmässige Audits zur Überprüfung der KI-Transparenz und Fairness.
	Validate	Validierung der KI-Generierungen durch Vergleich mit manuellen Ergebnissen.	Anpassung der institutionellen Prozesse an KI-Unterstützung.	Wissenschaftliche Evaluierung der Ergebnisse und wiederkehrende Prüfung.

Tabelle 21: Aufgaben und Verantwortlichkeiten der Akteurinnen und Akteure entlang des MLOps-Prozesses.

Die MLOps-Phasen verdeutlichen, dass eine erfolgreiche Implementierung nicht nur eine technologische, sondern auch eine organisatorische, ethische und regulatorische Herausforderung darstellt. Eine enge Zusammenarbeit zwischen Softwareunternehmen, Institutionen und externen Expertinnen und Experten ist daher entscheidend, um eine sichere und effiziente KI-Integration in der Sozialen Arbeit nachhaltig zu ermöglichen.

Da diese Berichte direkte Auswirkungen auf betreute Personen haben, müssen Datenschutz, Transparenz, Fairness und Verantwortlichkeit besonders beachtet werden. Daher werden die ethischen Herausforderungen im folgenden Kapitel nochmals beleuchtet.

10.3 Ethische Herausforderungen und Verantwortung

Die Einführung von KI-gestützter Berichterstellung in der Sozialen Arbeit wirft neben technischen auch tiefgreifende ethische Fragen auf. Da diese Berichte direkte Auswirkungen auf betreute Personen haben, müssen Datenschutz, Transparenz, Fairness und Verantwortlichkeit besonders beachtet werden. Wie in Kapitel 3.4 *Abgrenzung* erwähnt, kann keine fundierte Auseinandersetzung mit der Thematik im Rahmen dieser Arbeit stattfinden. Aufgrund der sensiblen Datengrundlage werden aber folgende Themen explizit erwähnt:

- Datenschutz und Datensouveränität**
 Wie in Kapitel 3.4 *Abgrenzung – Absatz Datenschutz* dargelegt, gelten für personenbezogene Daten der Sozialen Arbeit besonders strenge Vorschriften. Die Analyse in Kapitel 8.3 *Datenanalyse* zeigt zudem, dass keine standardisierten Mechanismen zur Anonymisierung existieren, wodurch sensible Informationen unzureichend geschützt sein könnten. Unklar ist zudem, wem die Daten gehören und wer sie zu welchem Zweck nutzen darf. Besonders kritisch ist, ob bestehende Daten für weiterführende Modelltrainings genutzt werden, um neue Produkte oder gar Personas zu generieren, die den ethischen Grundsätzen der Sozialen Arbeit widersprechen. Auch stellt sich die Frage, wie mit Fällen umzugehen ist, in denen Personen den Einsatz von KI ablehnen und ob somit ganze Systemketten nicht mehr genutzt werden können.
- Transparenz und Nachvollziehbarkeit**
 KI-Modelle treffen Entscheidungen auf Basis komplexer Mechanismen, die für Fachpersonen schwer nachvollziehbar sind. Wie in Kapitel 8.4 *Erstellung von Berichten* beschrieben, enthalten generierte Texte Halluzinationen, die schwer zu erkennen sind. Ein verantwortungsvoller Einsatz erfordert Mechanismen zur Nachvollziehbarkeit, etwa durch strukturierte Validierungsprozesse, die in Kapitel 10.2.2 *Institutionen* diskutiert wurden. Die hier aufgezeigte Praxis zeigt jedoch, dass KI-gestützte Berichterstellung in der Sozialen Arbeit noch am Anfang steht. Während andere Branchen standardisierte Methoden zur Erklärbarkeit entwickelt haben, fehlen hier klare Richtlinien.
- Verzerrungen**
 Wie in Kapitel 10.2.1 *GLAUX GROUP* beschrieben, könnten Ansätze zur Bias-Erkennung und Datenprüfung in die Software integriert werden. Dies allein reicht jedoch nicht aus - Fachpersonen müssen sich der möglichen Verzerrungen bewusst sein, sowohl durch technische Limitationen als auch durch eigene subjektive Einschätzungen. Offene Fragen bleiben, wie Verzerrungen in grossen Datensätzen zuverlässig erkannt und minimiert werden können.

- **Forschung**
Es wird davon ausgegangen, dass weitere Studien nötig sind, um zu bewerten, ob und unter welchen Bedingungen KI die Qualität der Berichterstellung tatsächlich verbessern kann und ob Modelle fundiert auf diese spezifische Domäne trainiert werden können. Da die Dokumentation häufig in Freitext erfolgt, kann dies die strukturierte Analyse und Modellanpassung erschweren.
- **Verantwortung und Haftung**
Fehlentscheidungen in der Sozialen Arbeit können schwerwiegende Folgen haben. Wenn KI fehlerhafte Inhalte produziert, liegt die Verantwortung weiterhin bei den Fachpersonen und Institutionen, da aktuelle KI-Systeme nicht autonom arbeiten. Damit bleibt die Frage der Verantwortung und Haftung offen.
- **Wirtschaftliche Interessen vs. ethische Verantwortung**
Obwohl KI-gestützte Lösungen für Softwareunternehmen finanzielle Anreize und Wettbewerbsvorteile schaffen, müssen ethische Grundsätze Vorrang haben. Fehlerhafte oder unzureichend validierte Berichte können schwerwiegende Folgen für betreute Personen haben. Institutionen tragen hierbei eine unmittelbare Verantwortung, doch auch Softwareunternehmen müssen sicherstellen, dass ihre Systeme hohe fachliche und ethische Standards erfüllen.
- **Mögliche Folgen von KI in der Sozialen Arbeit**
Der Einsatz von KI in der Sozialen Arbeit kann weitreichende Folgen haben. Automatisierte Berichte könnten das Vertrauen in die fachliche Einschätzung von Fachpersonen verringern, wenn KI-generierte Texte unkritisch übernommen werden. Zudem besteht das Risiko, dass bestehende soziale Ungleichheiten verstärkt werden, da KI-Modelle auf historischen Daten basieren und darin enthaltene Vorurteile reproduzieren können. Institutionen könnten KI als effiziente Lösung sehen und von Fachkräften erwarten, sich stärker darauf zu verlassen, selbst wenn die Ergebnisse nicht immer kontextsensitiv oder tiefgehend sind. Daher ist es essenziell, dass Fachpersonen ihre interpretative Rolle bewahren, Bias aktiv reflektiert wird und KI nicht als objektive Wahrheit, sondern als unterstützendes Werkzeug verstanden wird.

KI-gestützte Berichterstellung bietet Potenzial, bringt jedoch erhebliche ethische Herausforderungen mit sich. Datenschutz, Transparenz, Fairness und Verantwortung sind nicht nur technisch, sondern auch organisatorisch zu adressieren. Linnemann et al. betonen dabei, dass Algorithmen in der Sozialen Arbeit nur dann einen Mehrwert bieten, wenn Fachkräfte in die Entwicklung einbezogen werden und eine transparente Nutzung der Daten sichergestellt ist [6, S. 12]. Besonders kritisch ist, dass es für Fachpersonen schwierig wird, Verzerrungen zuverlässig zu erkennen und es an validen Methoden zur Nachvollziehbarkeit fehlt. Zudem sind noch viele Fragen offen – etwa zur Nutzung von Daten für weiterführende Modelltrainings oder zu den Folgen individueller Ablehnungen gegenüber KI.

10.4 Reflexion der methodischen Vorgehensweise

Die methodische Vorgehensweise dieser Arbeit war darauf ausgerichtet, praxisnahe Erkenntnisse zur KI-gestützten Berichterstellung in der Sozialen Arbeit zu gewinnen. Dabei wurde ein explorativer Ansatz gewählt, um sowohl technische als auch fachliche Herausforderungen zu identifizieren und Lösungsansätze für die Applikation *socialweb* abzuleiten. Trotz einer fundierten Analyse gibt es einige methodische Limitationen, die bei der Interpretation der Ergebnisse berücksichtigt werden müssen.

- **Begrenzte wissenschaftliche Fachliteratur**
Aufgrund der Schnelllebigkeit der Technologie und der relativ neuen Entwicklungen im Bereich der KI-gestützten Textgenerierung existiert nur eine begrenzte Anzahl an wissenschaftlichen Publikationen. Viele relevante Erkenntnisse stammen aus technischen Dokumentationen, Unternehmensberichten oder Preprint-Studien, die noch nicht peer-reviewed sind. Dies stellt eine methodische Einschränkung dar, da etablierte akademische Referenzen fehlen, auf die sich die Analyse stützen könnte.

- Begrenzung der Datenbasis und Repräsentativität**
 Die begrenzte Datenbasis stellt eine zentrale Einschränkung dar. Die Analyse von drei Dossiers innerhalb einer Institution bietet zwar erste Einblicke, erlaubt jedoch keine repräsentativen Aussagen für die gesamte Soziale Arbeit in der Schweiz. Unterschiede zwischen Institutionen, Fachbereichen oder kantonalen Vorgaben blieben unberücksichtigt, was die Übertragbarkeit der Ergebnisse einschränkt. Zudem besteht das Risiko, dass spezifische Muster überbewertet und andere relevante Faktoren vernachlässigt werden. Eine breitere Datengrundlage wäre notwendig, um Verzerrungen zu minimieren und strukturelle Unterschiede in Dokumentationspraktiken zu identifizieren. Dies erfordert jedoch erheblichen Aufwand, da Erfassungsmethoden harmonisiert und Vergleichskriterien definiert werden müssten. Künftige Studien sollten daher gezielt auf eine erweiterte Datenbasis ausgerichtet sein, um belastbare Erkenntnisse zu gewinnen.
- Einfluss der Datenqualität auf die Ergebnisse**
 Wie mehrfach dargestellt, spielt die Datenqualität eine zentrale Rolle für die Leistung KI-gestützter Berichtssysteme. Die vorhandenen Daten wiesen Inkonsistenzen, fehlende Strukturierung und inhaltliche Lücken auf, wodurch eine valide Bewertung potenzieller KI-Anwendungen erschwert wurde. Dies unterstreicht die Notwendigkeit einer verbesserten Datenvorbereitung und standardisierten Dokumentation, bevor weiterführende KI-Modelle getestet oder implementiert werden können. Die Aufwände für die fundierte Analyse der Daten wurden unterschätzt und müssen bei einer weiterführenden Prüfung adressiert werden.
- Subjektivität in der Analyse und Interpretation**
 Ein weiterer methodischer Aspekt betrifft die subjektive Interpretation der Daten. Die Einordnung und Bewertung der Herausforderungen basierte auf einer Kombination aus qualitativen Beobachtungen und einer theoretischen Herleitung. Zudem erfolgte die Evaluierung der Ergebnisse ausschliesslich durch den Autor, wodurch methodische Einschränkungen entstehen. Eine breitere Validierung durch weitere Fachpersonen wäre erforderlich gewesen, um eigene Verzerrungen zu minimieren und eine höhere Repräsentativität der Ergebnisse sicherzustellen. Dies ist insbesondere bei explorativen Ansätzen wichtig, da unterschiedliche Fachperspektiven zu einer fundierteren Einordnung der Resultate beitragen können.
- Einfluss der eingesetzten KI-Modelle**
 Da sich die Analyse auf ein spezifisches Sprachmodell konzentrierte, konnten alternative KI-Modelle nicht systematisch verglichen werden. Andere Modelle könnten möglicherweise eine bessere Leistung in der Sozialen Arbeit erzielen, insbesondere wenn sie gezielt für diesen Fachbereich trainiert würden. Eine umfassendere Untersuchung verschiedener Modelle wäre erforderlich, um fundierte Aussagen zur optimalen technischen Umsetzung zu treffen. Ausgehend von der Komplexität und den fehlenden Kenntnissen des Autors konnten die Modelle nicht fundiert lokal getestet werden. Eine weiterführende Untersuchung mit spezialisierten Modellen und ausgewählten IT-Fachpersonen wäre notwendig, um differenziertere Aussagen zur Modellqualität zu treffen.
- Zusammenarbeit mit Partnerinnen und Partner**
 Das Produkt *SwissGPT* konnte im Rahmen dieser Arbeit nicht validiert werden (siehe Kapitel 8.1.2 *Stossrichtung 2: SwissGPT von AlpineAI*). Dennoch bleibt die Zusammenarbeit mit spezialisierten Technologieunternehmen ein relevanter Aspekt, um praxisnahe Lösungen für die Soziale Arbeit zu entwickeln. Es wird empfohlen, solche Partnerschaften gezielt zu evaluieren, insbesondere im Hinblick auf Datenschutz, Sprachmodelle für den deutschsprachigen Raum und die Anpassbarkeit an fachliche Standards. Zudem könnten Kooperationen mit Forschungseinrichtungen helfen, KI-Modelle spezifisch für die Soziale Arbeit zu optimieren. Ein kontinuierlicher Austausch mit relevanten Beteiligten wäre notwendig, um technologische Entwicklungen frühzeitig einzubeziehen und mögliche Synergien zu nutzen.

- **Iterativer Charakter der Methodik**

Ein zentrales Merkmal der gewählten Methodik ist ihr iterativer Charakter. Die Analyse zeigte, dass viele Herausforderungen erst während des Forschungsprozesses deutlich wurden – insbesondere im Hinblick auf die Anforderungen an Datenqualität, Dokumentationsstandards und die Notwendigkeit menschlicher Validierung (siehe Kapitel 10.3 *Ethische Herausforderungen und Verantwortung*). Dies verdeutlicht, dass die Einführung KI-gestützter Berichterstellung nicht als einmaliges Implementierungsprojekt verstanden werden kann, sondern als fortlaufender Prozess, der regelmäßige Anpassungen und Evaluierungen erfordert.

- **Interdisziplinäre Forschungsteams**

Für die KI-gestützte Berichterstellung ist eine breitere interdisziplinäre Expertise erforderlich, die über einzelne Fachdisziplinen hinausgeht. Eine zukünftige Untersuchung sollte idealerweise durch ein Forschungsteam mit Expertinnen und Experten aus Sozialer Arbeit, Informatik, Ethik und Recht erfolgen. Dies würde ermöglichen, technische, methodische und ethische Fragestellungen gezielter zu analysieren und fundiertere Schlussfolgerungen für die Praxis abzuleiten.

Die methodische Reflexion zeigt, dass diese Arbeit einen ersten, explorativen Schritt darstellt, um KI in der Sozialen Arbeit systematisch zu untersuchen. Die identifizierten Limitationen verdeutlichen gleichzeitig, dass weitere Forschung und praktische Tests erforderlich sind, um eine verlässliche Implementierung zu ermöglichen.

10.5 Alternative KI-Ansätze in der Praxis

Neben der KI-gestützten Berichterstellung wurden weitere Ansätze evaluiert. Diese zeigen unterschiedliche Herausforderungen und Potenziale auf und liefern wertvolle Erkenntnisse für zukünftige Implementierungen in der Sozialen Arbeit. Während ein Ansatz den Fokus auf die Verarbeitung strukturierter Daten legt, wurde zusätzlich untersucht, wie die Integration von KI in *socialweb* bestehende Workflows bei der Arbeitsplanung verbessern kann. Die daraus gewonnenen Erkenntnisse ermöglichen ein besseres Verständnis technischer und organisatorischer Anforderungen und helfen, gezielte Optimierungen für den praktischen Einsatz abzuleiten.

10.5.1 Handbuch-Chatbot

Neben dieser Masterarbeit wurde ein Chatbot für ein anderes Produkt der *GLAUX GROUP* evaluiert, der Mitarbeitenden den Zugriff auf das bestehende Handbuch erleichtern sollte. Hierzu wurde ein bestehender Technologiestack genutzt, auf dessen Basis die Inhalte des Handbuchs eingespeist und für die Interaktion mit einem LLM aufbereitet wurden. Ziel war es, den Nutzerinnen und Nutzern zu ermöglichen, gezielt Fragen zum Produkt zu stellen und präzise, kontextbezogene Antworten zu erhalten.

Die Resultate wurden im Technical Board der *GLAUX GROUP* präsentiert:

- Die ersten Tests mit dem externen Partner lieferten nicht die gewünschten Ergebnisse. Trotz der strukturierten Eingabe der Handbuchinhalte waren die generierten Antworten teilweise ungenau oder enthielten Halluzinationen. Besonders problematisch war, dass die Verzerrungen nicht immer sofort erkennbar waren, was das Risiko fehlerhafter Informationen erhöhte.
- Parallel dazu wurden interne Tests mit alternativen Modellen durchgeführt, die bessere, aber dennoch nicht zufriedenstellende Ergebnisse lieferten. Dies verdeutlicht, dass selbst mit einer klar definierten Wissensbasis Herausforderungen in der Verarbeitung natürlicher Sprache mit NLP bestehen.

Obwohl dieses Projekt thematisch von der KI-gestützten Berichterstellung abweicht, zeigt es doch ähnliche Herausforderungen im Umgang mit LLMs. Die Schwierigkeit, zuverlässige und fehlerfreie Antworten aus bestehenden Dokumentationen zu generieren, ist ein zentrales Problem, das auch bei der automatisierten Erstellung von Berichten berücksichtigt werden muss. Da die Ergebnisse nicht vertieft ausgewertet wurden, sind keine fundierten Rückschlüsse auf spezifische Optimierungsmöglichkeiten

möglich. Dennoch unterstreicht dieses Experiment die Notwendigkeit, KI-generierte Inhalte kritisch zu hinterfragen und durch ergänzende Validierungsmechanismen abzusichern.

10.5.2 Automatisierte Arbeitsplanung

Zusätzlich zu dieser Masterarbeit wurde ein weiteres KI-Projekt durch den Autor innerhalb von *socialweb* angestoßen, um den Einsatz von KI gezielt weiterzuentwickeln. Während sich diese Arbeit auf die KI-gestützte Berichterstellung konzentrierte, wurde parallel die Automatisierung der Arbeitsplanung als eigenständige Initiative angestoßen. Diese Thematik wurde aufgrund von Rückmeldungen aus den Institutionen als besonders relevant identifiziert, da die Erstellung von Dienstplänen, insbesondere in grossen Organisationen, eine erhebliche Herausforderung darstellt. Neben personellen Verfügbarkeiten müssen zahlreiche weitere Faktoren wie Qualifikationen, gesetzliche Vorgaben, individuelle Wünsche und betriebliche Notwendigkeiten berücksichtigt werden. Die aktuelle manuelle Planung ist zeitaufwendig, fehleranfällig und oft ineffizient.

Im Gegensatz zur KI-gestützten Berichterstellung, bei der unstrukturierte Textdaten mit NLP verarbeitet werden müssen, bietet das Modul *Planung* in *socialweb* entscheidende Vorteile: Die vorhandenen Daten sind strukturiert, standardisiert und anonymisiert, was eine wesentlich zuverlässigere Modellierung ermöglicht. Zudem existieren bereits leistungsfähige mathematische Optimierungsverfahren und sogenannte Solver, die speziell für die automatische Einsatz- und Schichtplanung entwickelt wurden. Dies reduziert den Entwicklungsaufwand erheblich und erhöht die Erfolgchancen einer KI-gestützten Lösung.

Vor diesem Hintergrund wurde im Dezember 2024 ein PoC durchgeführt, um die Machbarkeit einer automatisierten Planung zu evaluieren. Dabei wurden bestehende Datenstrukturen genutzt und ein auf ML basierender Ansatz getestet. Erste Ergebnisse (siehe nachstehende Grafik) zeigten sehr positive Resultate und es wurde entschieden, die Implementierung weiter voranzutreiben. Derzeit wird ein lokaler Prototyp entwickelt, der zunächst in einer begrenzten Testumgebung validiert wird. Dabei sollen Erfahrungswerte aus der Praxis in die Weiterentwicklung einfließen, um sicherzustellen, dass das System nicht nur technisch, sondern auch organisatorisch optimal integriert werden kann.

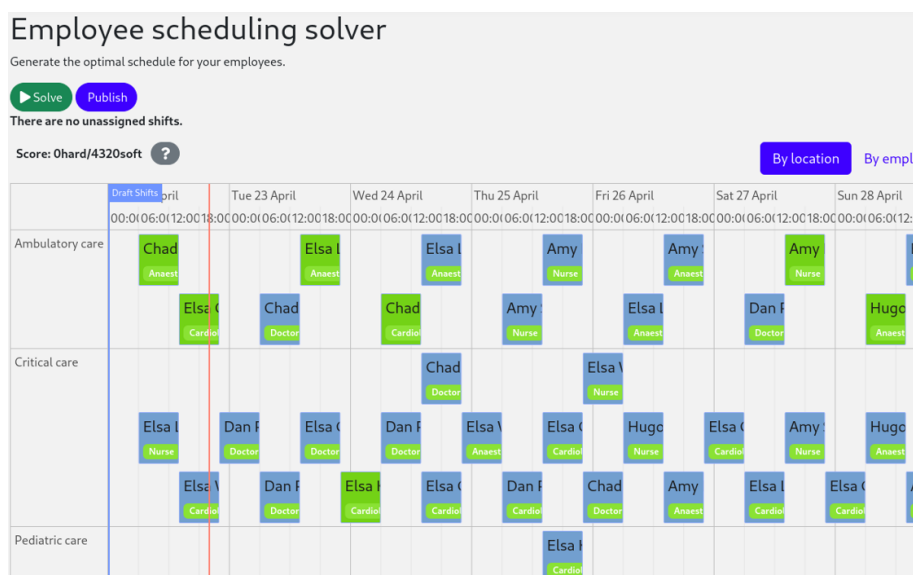


Abbildung 39: Printsreen *PoC Automatisierte Arbeitsplanung* in *socialweb*.

Die Beta-Version des Systems ist für das zweite Quartal 2025 geplant. Sollte sich der Prototyp bewähren, könnte dies einen wesentlichen Schritt zur Optimierung der Arbeitsplanung in *socialweb* darstellen. Neben einer verbesserten Planungseffizienz könnten durch den Einsatz von KI langfristig auch administrative Ressourcen eingespart und eine gerechtere Verteilung der Arbeitszeiten ermöglicht werden.

11 Schlussfolgerungen, Empfehlungen und Fazit

Die vorliegende Masterarbeit untersuchte die Möglichkeiten und Herausforderungen der KI-gestützten Berichterstellung in der Sozialen Arbeit, insbesondere im Kontext der Applikation *socialweb*. Während die technologische Machbarkeit grundsätzlich gegeben ist, zeigen sich wesentliche Einschränkungen hinsichtlich Datenqualität, Transparenz, Reproduzierbarkeit, ethischer Fragestellungen und einer unzureichenden Ergebnisqualität.

Dieses Schlusskapitel fasst die zentralen Erkenntnisse zusammen und beantwortet die Haupt- und Teilforschungsfragen auf technischer, sozialwissenschaftlicher und wirtschaftlicher Ebene. Zudem werden Handlungsempfehlungen für Institutionen und die *GLAUX GROUP* abgeleitet, um eine praxisnahe Integration von KI in die Dokumentationsprozesse der Sozialen Arbeit zu ermöglichen.

11.1 Beantwortung der Hauptforschungsfrage

Welche technischen, methodischen und organisatorischen Voraussetzungen sind erforderlich, um KI in der Softwarelösung *socialweb* für die automatisierte Berichterstellung zu implementieren?

Die Untersuchung zeigt, dass eine KI-basierte Berichterstellung technisch realisierbar ist, jedoch mit wesentlichen Einschränkungen verbunden bleibt. Während Sprachmodelle wie *GPT-4o* von *OpenAI* in der Lage sind, Texte mit hoher sprachlicher Qualität zu generieren, bestehen erhebliche Herausforderungen in Bezug auf Halluzinationen, mangelnde Reproduzierbarkeit und fehlende Transparenz. Die Qualität der generierten Berichte hängt dabei massgeblich von der Konsistenz und Struktur der zugrunde liegenden Daten ab. Eine unzureichende Datenbasis kann dazu führen, dass die KI zentrale Informationen nicht erkennt oder falsch interpretiert. Daher ist eine manuelle Überprüfung durch Fachpersonen unverzichtbar, um Fehlschlüsse und unvollständige Inhalte zu vermeiden.

Neben den technischen Aspekten erfordert die Implementierung einer KI-gestützten Berichterstellung auch methodische Anpassungen. Eine strukturierte und standardisierte Erfassung der Daten ist entscheidend, um die Qualität der generierten Berichte zu gewährleisten. Zudem müssen Mechanismen zur Identifikation und Reduktion von Verzerrungen integriert werden, um eine objektive und faire Berichterstattung sicherzustellen. Die KI darf nicht als vollständig autonome Lösung betrachtet werden, sondern sollte als Assistenzsystem fungieren, das Fachpersonen durch automatisierte Vorschläge und strukturierte Textbausteine unterstützt.

Auf organisatorischer Ebene sind weitere Voraussetzungen notwendig, um eine erfolgreiche Implementierung in der Praxis zu ermöglichen. Fachpersonen müssen im Umgang mit KI-gestützten Systemen geschult werden, um generierte Inhalte kritisch zu bewerten und zu validieren. Zudem bedarf es einer engen interdisziplinären Zusammenarbeit zwischen Expertinnen und Experten aus der IT, der Sozialen Arbeit und der Ethik, um sicherzustellen, dass sowohl technische als auch ethische Anforderungen erfüllt werden. Dabei muss ein besonderer Fokus auf den Datenschutz gelegt werden, um sensible Informationen zu schützen und gesetzliche Vorgaben einzuhalten. Darüber hinaus sind klare Prozesse zur Qualitätssicherung und Nachbearbeitung der Berichte erforderlich, um die Verlässlichkeit der KI-generierten Inhalte zu gewährleisten.

Insgesamt verdeutlichen die Ergebnisse, dass eine direkte und vollautomatische Integration von KI in die Berichterstellung innerhalb der Softwarelösung *socialweb* derzeit nicht umsetzbar ist. Allerdings bietet KI grosses Potenzial als unterstützendes Werkzeug, insbesondere durch automatisierte Vorschläge zur Strukturierung und Formulierung von Berichten, die von Fachpersonen überprüft und angepasst werden. Damit dieses Potenzial voll ausgeschöpft werden kann, müssen die technischen, methodischen und organisatorischen Voraussetzungen gezielt weiterentwickelt und optimiert werden.

Eine detaillierte Betrachtung dieser Aspekte findet sich in den Kapiteln der untenstehenden Tabelle.

Themen	Relevante Kapitel
Qualitätssicherung, Halluzinationen, Bias	4.2.2 Trainingsdaten und ihre Implikationen, 4.2.3 Prompting, 4.2.5 Evaluierungsmethoden, 8.4 Erstellung von Berichten
Datenqualität	4.2.2 Trainingsdaten und ihre Implikationen, 4.2.4 Qualität der Eingabedaten, 10.1 Einfluss der Datenqualität auf MLOps
Evaluierung	8.4.4 Evaluierung der Ergebnisqualität
Optimierungsmöglichkeiten	10.2 Akteurinnen und Akteure im MLOps-Prozess

Tabelle 22: Themen und relevante Kapitel für die Beantwortung der Hauptforschungsfrage.

11.2 Beantwortung der technischen Forschungsfragen

Welche technischen Anforderungen müssen erfüllt sein, um eine KI-basierte Berichterstellung in *socialweb* zu implementieren und gleichzeitig die datenschutzrechtlichen Vorgaben zu wahren?

Die Untersuchung zeigt, dass Datenanonymisierung zwingend erforderlich ist, wenn ein öffentlich zugängliches Modell verwendet wird. Bei einer lokalen Implementierung muss der Anonymisierungsgrad individuell geprüft werden. In diesem Fall muss berücksichtigt werden, dass während der Analyse, des Modelltrainings und der Evaluation sensible Daten potenziell von vielen Fachpersonen, darunter auch IT-Spezialistinnen und IT-Spezialisten, eingesehen werden.

Eine strukturierte Datenaufbereitung und ein standardisiertes Prompting sind essenziell, um konsistente Eingaben sicherzustellen und die Ergebnisqualität zu optimieren. Zur Verbesserung der Transparenz könnte ein kontrollierbares Modell, etwa mittels RAG, eingesetzt werden. Dieses Verfahren wurde jedoch nicht weiter untersucht.

Themen	Relevante Kapitel
Datenschutz	3.4 Abgrenzung (Datenschutz), 4.3.2 Implementationsverfahren, 6.2 Datenbeschaffung, Analyse und Vorbereitung, 8.3 Datenanalyse
Transparenz und Nachvollziehbarkeit	4.2.3 Prompting, 8.4 Erstellung von Berichten, 8.5 Zusammenfassung und Fazit
Technische Umsetzung und Architektur	4 Stand der Forschung Teil 1 – Technologische Perspektiven, 8.1 Proof of Concept und Bestimmung der Stossrichtung, 7 Methoden Teil 2 – Wirtschaftliche Perspektiven

Tabelle 23: Themen und relevante Kapitel für die Beantwortung der 1. technischen Frage.

Welche methodischen Anpassungen in der Datenerfassung und Datenaufbereitung werden benötigt?

Die Software stellt bereits verschiedene Möglichkeiten zur validen Datenerfassung bereit, jedoch zeigt sich, dass diese im Alltag der Fachpersonen nicht konsequent genutzt werden. Dies führt zu Inkonsistenzen in der Datenqualität, die sich direkt auf die Verlässlichkeit der KI-generierten Berichte auswirken. Um dies zu verbessern, braucht es klare methodische Vorgaben seitens der Institutionen, die in Zusammenarbeit mit externen Akteurinnen und Akteuren zielgerichtete Standards für die Datenerfassung definieren. Gleichzeitig kann die *GLAUX GROUP* weitere technische Möglichkeiten evaluieren, um Fachpersonen im Arbeitsalltag besser zu unterstützen und die Nutzung der bestehenden Funktionen zu optimieren.

Themen	Relevante Kapitel
Strukturierte Datenerfassung	4.2.4 Qualität der Eingabedaten, 5.3 Dokumentation und Berichtswesen, 8.3 Datenanalyse
Optimierung der Datenqualität	10.2 Akteurinnen und Akteure im MLOps-Prozess

Tabelle 24: Themen und relevante Kapitel für die Beantwortung der 2. technischen Frage.

Ist eine Datenanonymisierung zwingend erforderlich, um KI datenschutzkonform in *socialweb* zu integrieren?

Die Untersuchung zeigt, dass eine Datenanonymisierung zwingend erforderlich ist, wenn ein externes, öffentlich zugängliches Modell genutzt wird. In diesem Fall müssen personenbezogene Daten vor der Verarbeitung durch die KI entfernt oder durch geeignete Anonymisierungsverfahren geschützt werden. Wird das Modell hingegen auf einer dedizierten, lokal betriebenen Infrastruktur eingesetzt, ist eine vollständige Anonymisierung nicht zwingend erforderlich. Entscheidend ist hierbei, dass sämtliche Daten innerhalb der Institution verarbeitet werden und klare Zugriffsbeschränkungen bestehen. Ein alternativer Ansatz zur Wahrung der Datensicherheit ist RAG. Diese Methode ermöglicht es, sensible Informationen extern zu speichern und dennoch kontextspezifisch für KI-Anwendungen zu nutzen, ohne die Privatsphäre der betroffenen Personen zu gefährden. Eine vertiefte Untersuchung dieses Verfahrens war jedoch nicht Bestandteil dieser Arbeit.

Themen	Relevante Kapitel
Datenanonymisierung	4.1.8 Retrieval Augmented Generation, 4.3.2 Implementationsverfahren
Lokale vs. externe Verarbeitung	4.3.2 Implementationsverfahren
Datenschutzkonforme KI-Nutzung	4.1.8 Retrieval Augmented Generation, 4.2.7 Explainable AI

Tabelle 25: Themen und relevante Kapitel für die Beantwortung der 3. technischen Frage.

Wie lässt sich die Ergebnisqualität der KI-generierten Berichte bewerten und kann diese durch Techniken wie Prompting optimiert werden?

Die Bewertung der Ergebnisqualität der KI-generierten Berichte kann sowohl durch automatisierte als auch manuelle Methoden erfolgen. Automatisierte Verfahren, wie beispielsweise statistische Analysen oder Vergleichstests mit Referenztexten, bieten zwar eine erste Einschätzung, sind jedoch oft nicht ausreichend fundiert implementiert. Dies führt dazu, dass qualitative Mängel, etwa fehlerhafte Interpretationen oder kontextuelle Verzerrungen, nicht zuverlässig erkannt werden. Daher ist eine manuelle Evaluierung durch Fachpersonen unerlässlich, um die inhaltliche Korrektheit, Vollständigkeit und fachliche Relevanz der Berichte sicherzustellen.

Die Optimierung der Ergebnisqualität durch Prompting ist möglich, jedoch nicht vollständig steuerbar. Es gibt nicht den einen perfekten Prompt, sondern bestenfalls eine iterative Interaktion zwischen der Fachperson und der KI. Durch wiederholtes Anpassen der Eingaben können gezielt präzisere und relevantere Berichte generiert werden. Dennoch bleibt die Variabilität der KI-Ausgaben eine Herausforderung, weshalb strukturierte Prompts zwar hilfreich, aber keine Garantie für konsistente und fehlerfreie Ergebnisse sind.

Themen	Relevante Kapitel
Evaluierung und Qualitätssicherung	4.2.5 Evaluierungsmethoden, 6.6 Evaluierung, 8.4 Erstellung von Berichten, 10.2 Akteurinnen und Akteure im MLOps-Prozess
Optimierung durch Prompting	4.2.3 Prompting, 8.4.2 Ergänzendes und verbessertes Prompting

Tabelle 26: Themen und relevante Kapitel für die Beantwortung der 4. technischen Frage.

Wie entstehen Verzerrungen (Bias) und mit welchen Methoden lassen sich diese minimieren?

Eine hohe Datenqualität ist entscheidend, um Verzerrungen (Bias) in KI-generierten Berichten zu minimieren. Zudem führen unvollständige, unausgewogene oder fehlerhafte Trainingsdaten dazu, dass das Modell Muster übernimmt, die nicht repräsentativ oder sachlich nicht korrekt sind. Besonders problematisch sind Daten, die bestimmte Perspektiven überbetonen oder wichtige Informationen unzureichend abbilden.

Die Minimierung von Bias erfordert einen iterativen Prozess auf zwei Ebenen. Erstens ist eine kontinuierliche Interaktion zwischen Fachpersonen und KI notwendig, um durch gezielte Anpassungen in der Dateneingabe und der Interpretation der Ergebnisse Verzerrungen frühzeitig zu erkennen. Zweitens müssen gemäss MLOps-Prinzipien nach Farah Modelle regelmässig überprüft, nachtrainiert und validiert werden. Eine gezielte Auswahl und Nachbearbeitung der Input- und Trainingsdaten, kombiniert mit einer systematischen Evaluierung der Modellleistung, ist essenziell, um Verzerrungen nachhaltig zu reduzieren und eine möglichst objektive Berichterstellung zu gewährleisten.

Themen	Relevante Kapitel
Bias-Entstehung	4.2.2 Trainingsdaten und ihre Implikationen, 4.2.3 Prompting, 4.2.4 Qualität der Eingabedaten 4.2.4 Qualität der Eingabedaten
Methoden zur Bias-Reduktion	4.2.3 Prompting, 6.5 Bias-Analyse und Reduktion, 8.4.3 Bias-Analyse und Reduktion
Fachliche Kontrolle	10.3 Ethische Herausforderungen und Verantwortung

Tabelle 27: Themen und relevante Kapitel für die Beantwortung der 5. technischen Frage.

11.3 Beantwortung der Forschungsfragen im Zusammenhang mit der Sozialen Arbeit

Welche spezifischen Herausforderungen und Anforderungen ergeben sich in der Sozialen Arbeit bei der Dokumentation und der Erstellung von Berichten?

Die Grundlage für das Erstellen von Berichten in der Sozialen Arbeit sind mehrschichtige Dokumentationsprozesse, die oft mit einem hohen Aufwand verbunden sind und Fachpersonen stark beanspruchen. Die Zeit, die für die Erstellung von Berichten aufgewendet wird, fehlt häufig für die direkte Arbeit mit Klientinnen und Klienten. Gleichzeitig müssen die Berichte nicht nur präzise und rechtlich korrekt sein, sondern auch eine fachlich fundierte Reflexion enthalten. Dies führt zu einem Spannungsfeld zwischen Effizienz, fachlicher und inhaltlicher Tiefe.

Eine weitere Herausforderung ist die unklare Zielsetzung der Datenerfassung. In der Praxis zeigt sich, dass Dokumentationen neben ihrem eigentlichen Zweck häufig auch interne Themen beinhalten, etwa die Verteilung von Aufgaben innerhalb des Teams. Dies führt zu uneinheitlichen Erfassungsmethoden und erschwert eine strukturierte Nutzung der Daten. Um dem entgegenzuwirken, braucht es verbindliche Standards, die von den Institutionen definiert und durch übergeordnete Organe wie Politik, Hochschulen, Fachverbände, Berufsverbände sowie Aufsichts- und Sozialbehörden einheitlich vorgegeben werden. Nur durch eine klare Definition der Anforderungen kann sichergestellt werden, dass Dokumentationen ihren eigentlichen Zweck erfüllen und gleichzeitig eine effiziente sowie qualitativ hochwertige Berichterstellung ermöglicht wird.

Themen	Relevante Kapitel
Dokumentationsaufwand	5.3 Dokumentation und Berichtswesen
Standardisierung von Berichten	5.3 Dokumentation und Berichtswesen, 10.2.2 Institutionen
Datenschutz	3.4 Abgrenzung (Abschnitt Datenschutz), 4.3.2 Implementationsverfahren, 6.2 Datenbeschaffung, Analyse und Vorbereitung, 8.3 Datenanalyse

Tabelle 28: Themen und relevante Kapitel für die Beantwortung der 1. Frage im Zusammenhang mit der Sozialen Arbeit.

Welche ethischen Anforderungen und Verantwortlichkeiten müssen KI-Systeme in der Sozialen Arbeit erfüllen, um professionellen und sozialen Ansprüchen gerecht zu werden?

Der Einsatz von KI-Systemen in der Sozialen Arbeit muss hohen ethischen Anforderungen entsprechen, da er direkte Auswirkungen auf betreute Personen hat. Datenschutz, Transparenz, Fairness und Verantwortlichkeit sind dabei zentral.

Datenschutz und Datensouveränität sind unabdingbar, da personenbezogene Daten strengen Vorschriften unterliegen. In der Schweiz gehören die erfassten Daten den Klientinnen und Klienten, wodurch ihre Nutzung für Modelltrainings ohne ausdrückliche Einwilligung nicht zulässig ist. Dem Autor sind während der Arbeit keine standardisierten Anonymisierungsmechanismen bekannt geworden, die eine datenschutzkonforme Nutzung von KI-Systemen ohne Einschränkungen ermöglichen.

Transparenz und Nachvollziehbarkeit sind notwendig, da KI-Modelle komplexe Entscheidungen treffen, die Fachpersonen nicht immer nachvollziehen können. Validierungsprozesse müssen sicherstellen, dass KI-generierte Inhalte überprüft und korrigiert werden können.

Ein weiteres Risiko ist die De-Professionalisierung der Fachpersonen. KI darf menschliche Urteilsfähigkeit nicht ersetzen, da Soziale Arbeit Empathie, ethische Abwägungen und individuelle Fallbearbeitung erfordert. KI sollte Fachkräfte nur unterstützen, nicht ihre Kompetenz untergraben.

Auch die Verantwortung und Haftung bleiben kritisch. Fehlerhafte Berichte liegen in der Verantwortung der Fachpersonen, doch es bleibt offen, ob auch Softwareunternehmen bei negativen Folgen haftbar gemacht werden sollten. Klare regulatorische Leitlinien sind hier erforderlich.

Themen	Relevante Kapitel
Verantwortung und Ethik	5.2 Verantwortung und Ethik in der Anwendung, 10.3 Ethische Herausforderungen und Verantwortung
Transparenz und Nachvollziehbarkeit	4.2.3 Prompting, 10.3 Ethische Herausforderungen und Verantwortung
Einfluss auf Fachpersonen	5.2 Verantwortung und Ethik in der Anwendung, 5.3 Dokumentation und Berichtswesen

Tabelle 29: Themen und relevante Kapitel für die Beantwortung der 2. Frage im Zusammenhang mit der Sozialen Arbeit.

Wie und in welchen Bereichen wird KI bereits heute in der Sozialen Arbeit eingesetzt?

KI wird in der Sozialen Arbeit bereits in mehreren Bereichen eingesetzt, insbesondere zur Datenanalyse, als Entscheidungsunterstützung und zur Optimierung administrativer Prozesse. Die folgenden Beispiele zeigen einige zentrale Anwendungsfelder, ohne dass sie abschliessend sind.

- Ein Bereich ist das *Predictive Risk Modelling*, das Fachpersonen bei der Risikoabschätzung unterstützt, etwa durch das *Allegheny Family Screening Tool* in den USA. Erste Studien zeigen, dass solche Modelle in bestimmten Fällen präzise Einschätzungen liefern können.
- Chatbots und digitale Assistenzsysteme erleichtern den Zugang zu sozialen Dienstleistungen und entlasten Fachkräfte, indem sie Erstberatungen oder allgemeine Anfragen übernehmen.

NLP hilft bei der Analyse von Gesprächsprotokollen und Falldokumentationen, indem relevante Informationen strukturiert und schneller auffindbar gemacht werden.

- Zudem wird KI im Wissensmanagement genutzt, beispielsweise zur Kategorisierung von Beratungsprotokollen oder zur Vernetzung von Fachwissen.

Die Technologien bieten viel Potenzial, erfordern jedoch eine kontinuierliche Weiterentwicklung und sorgfältige Evaluierung, um den professionellen Anforderungen gerecht zu werden.

Themen	Relevante Kapitel
Bestehende KI-Anwendungen	5.1 Anwendungsmöglichkeiten von KI in der Sozialen Arbeit
Unterstützung von Fachpersonen	5.3 Dokumentation und Berichtswesen

Tabelle 30: Themen und relevante Kapitel für die Beantwortung der 3. Frage im Zusammenhang mit der Sozialen Arbeit.

11.4 Beantwortung der wirtschaftlichen Forschungsfragen

Die wirtschaftliche Bewertung einer KI-gestützten Berichterstellung in *socialweb* ist zum jetzigen Zeitpunkt nicht belastbar, da zentrale technische und organisatorische Voraussetzungen fehlen. Die nächsten Schritte müssen darauf abzielen, eine strukturierte Datenbasis zu schaffen und klare technische Rahmenbedingungen für eine zukünftige wirtschaftliche Analyse zu definieren. Entsprechend konnten folgende Forschungsfragen nicht geklärt werden:

- Welche wirtschaftlichen Aufwände sind erforderlich, um KI in *socialweb* zu integrieren und nachhaltig zu betreiben?
- Welches Kosten-Nutzen-Verhältnis ergibt sich durch die Implementierung von KI zur automatisierten Berichterstellung im Vergleich zur manuellen Erstellung?
- Wie hoch ist der ökonomische und funktionale Nutzen für die Institutionen und welcher Mehrwert ergibt sich daraus für die *GLAUX GROUP*?

Diese Fragen können erst beantwortet werden, wenn eine verlässliche Datenqualität und klare Rahmenbedingungen für die technische Implementierung geschaffen wurden.

Themen	Relevante Kapitel
Wirtschaftliche Unsicherheiten	9 Wirtschaftliche Ergebnisse
Datenqualität	4.2.2 Trainingsdaten und ihre Implikationen, 4.2.3 Prompting, 4.2.4 Qualität der Eingabedaten, 4.2.4 Qualität der Eingabedaten, 10.1 Einfluss der Datenqualität auf

Tabelle 31: Themen und relevante Kapitel für die Beantwortung der wirtschaftlichen Fragen.

11.5 Handlungsempfehlungen für die Praxis

Die Integration von KI-gestützter Berichterstellung in *socialweb* erfordert nicht nur technische Anpassungen, sondern auch strukturelle und organisatorische Massnahmen. Damit KI als unterstützendes Werkzeug erfolgreich eingesetzt werden kann, sind folgende Handlungsempfehlungen essenziell:

Handlungsempfehlungen für die *GLAUX GROUP*

- **Technische Funktionen erweitern**
Die *GLAUX GROUP* kann die bestehenden *socialweb*-Funktionen erweitern, um die Verarbeitung strukturierter Daten für KI-Anwendungen zu verbessern. Dazu gehört die Möglichkeit, Daten semantisch zu verknüpfen und KI-Modelle effizient in bestehende Workflows zu integrieren. Die Implementierung von KI-gestützten Prüfmechanismen in *socialweb* könnte helfen, Inkonsistenzen und Fehler in der Datenerfassung frühzeitig zu erkennen. Automatische Prüfungen bei der Eingabe könnten fachliche und sprachliche Validierungen in Echtzeit durchführen.
- **Schulungsangebote für Fachpersonen entwickeln**
Neben der technischen Umsetzung ist es essenziell, Fachpersonen gezielt im Umgang mit KI zu schulen. Die *GLAUX GROUP* könnte Fachschulung anbieten, um den sicheren und effizienten Einsatz der KI-gestützten Berichterstellung zu fördern.
- **Iteratives Vorgehen mit Pilotprojekten umsetzen**
Vor einer breiten Implementierung sollte ein PoC in enger Zusammenarbeit mit ausgewählten Institutionen durchgeführt und die Methoden dieser Masterarbeit validiert werden. So können Anforderungen frühzeitig evaluiert, Datenschutzaspekte geklärt und fachliche sowie technische Herausforderungen identifiziert werden.
- **Strategische Ausrichtung für zukünftige KI-Innovationen**
Neben der Berichterstellung bietet KI weiteres Potenzial in der Sozialen Arbeit. Die *GLAUX GROUP* sollte gezielt prüfen, welche Bereiche davon profitieren können, um Innovationen in *socialweb* voranzutreiben. Entscheidend ist dabei die Datenstruktur: KI arbeitet am zuverlässigsten mit strukturierten Daten, weshalb die Entwicklung auf solche Use Cases fokussiert werden sollte. Ein Beispiel ist der PoC zur automatisierten Arbeitsplanung, der im Dezember 2024 durchgeführt wurde. Erste Ergebnisse zeigen, dass ML-gestützte Optimierungsverfahren vielversprechend sind und schnell Nutzen für Institutionen bringen könnten.

Handlungsempfehlungen für Institutionen

- **Datenqualität als zentrale Voraussetzung verbessern**
Institutionen sollen verbindliche Standards für die Dokumentation entlang der bestehenden Fachkonzepte entwickeln, um eine konsistente und strukturierte Datenerfassung sicherzustellen. Dazu gehört die einheitliche Nutzung bestehender Funktionen in *socialweb* (z. B. Kategorien, Markierungen, Textbausteine) sowie regelmässige Schulungen für Fachpersonen, um die Qualität der Eingaben zu verbessern und zu überprüfen.

Auch manuell erstellte Berichte müssen validiert werden, da Fehler oder Inkonsistenzen in der Dokumentation die KI-gestützte Verarbeitung beeinträchtigen. Dies erfordert klare Qualitätssicherungsprozesse zur fachlichen Überprüfung von KI-generierten Texten.
- **IT-Schulungen für Fachpersonen und Management**
Um die erfolgreiche Umsetzung sowohl bestehender als auch neuer Massnahmen sicherzustellen, sollten sowohl Fachpersonen als auch das Management gezielt in IT-Kompetenzen geschult werden. Dies umfasst Grundlagen zu KI, Datenschutzbestimmungen sowie den praktischen Umgang mit digitalen Systemen.

- **Interdisziplinäre Zusammenarbeit fördern**

Die Einführung von KI erfordert eine enge Abstimmung zwischen Fachpersonen der Sozialen Arbeit, IT und Softwareunternehmen. Institutionen sollten gemeinsam mit Fachverbänden, Hochschulen und Datenschutzbehörden praxisnahe Standards für den KI-Einsatz definieren.

Diese Massnahmen legen die Grundlage für eine nachhaltige Implementierung von KI in der Sozialen Arbeit. Sie ermöglichen es Institutionen, die Potenziale der Technologie zu nutzen, ohne fachliche, ethische oder rechtliche Risiken einzugehen. Die erfolgreiche Implementierung von KI ist eine interdisziplinäre Aufgabe, die eine enge Zusammenarbeit zwischen Fachpersonen der Sozialen Arbeit, IT-Spezialistinnen und -Spezialisten sowie Entscheidungsträgerinnen und Entscheidungsträgern erfordert. Nur durch diesen integrativen Ansatz kann sichergestellt werden, dass KI technisch leistungsfähig, fachlich fundiert und ethisch vertretbar in bestehende Arbeitsprozesse integriert wird.

11.6 Ausblick

Die vorliegende Arbeit hat gezeigt, dass KI-gestützte Berichterstellung in der Sozialen Arbeit technisch möglich ist, jedoch mit erheblichen Herausforderungen verbunden bleibt. Neben der Datenqualität, Transparenz und ethischen Fragestellungen müssen auch die organisatorischen Rahmenbedingungen berücksichtigt werden. Damit KI langfristig einen echten Mehrwert bietet, sind weitere Forschungsarbeiten, technologische Weiterentwicklungen und praxisnahe Erprobungen erforderlich.

Die erfolgreiche Implementierung von KI in der Sozialen Arbeit erfordert eine enge Zusammenarbeit aller Beteiligten. Während die *GLAUX GROUP* die technologische Entwicklung vorantreibt, sind Institutionen und externe Akteurinnen und Akteure für die Gestaltung der organisatorischen, ethischen und rechtlichen Rahmenbedingungen verantwortlich. Die folgende Roadmap gibt einen Überblick über die geplanten Schritte aller Beteiligten in den kommenden Jahren:

Jahr	GLAUX GROUP	Institutionen	Externe
2025	Implementierung von KI mit strukturierten Daten im Modul <i>Planung</i>. Unterstützung der Institutionen bei der Erarbeitung fachlicher Standards in Zusammenarbeit mit anderen Akteurinnen und Akteuren. Ausarbeitung erster KI-Schulungsangebote.	Datenanalyse und Erfassung der aktuellen Dokumentationslage. Prüfung und Optimierung bestehender Dokumentationspraktiken. Sensibilisierung für KI in der Sozialen Arbeit, erste interne Diskussionen. Kontaktaufnahme zu übergeordneten Instanzen.	Analyse der regulatorischen, ethischen und datenschutzrechtlichen Rahmenbedingungen für den Einsatz von KI in der Sozialen Arbeit. Bereitstellung von Leitlinien und Empfehlungen für die Erarbeitung fachlicher Standards. Sensibilisierung von Politik und Gesellschaft für Chancen und Herausforderungen KI-gestützter Funktionen.
2026	Auswertung der KI-Funktionen im Modul <i>Planung</i>. Erstellung eines PoC für die KI-gestützte Datenerfassung im Modul <i>Journal</i>. Erstellung eines Golden Data Sets mit ausgewählten Institutionen. Identifikation optimaler KI-Modelle. Erhebung weiterer KI-Anforderungen in Zusammenarbeit mit Institutionen. Aufbau eines internen KI-Teams oder Kooperation mit spezialisierten Unternehmen	Aktive Zusammenarbeit mit GLAUX GROUP zur Evaluierung der ersten KI-Funktionen im Modul <i>Planung</i>. Validierung und Optimierung der Dokumentationspraktiken zur Sicherstellung einheitlicher und qualitativ hochwertiger Daten. Definition konkreter Anforderungen an die KI-gestützte Datenerfassung im Modul <i>Journal</i>. Unterstützung bei der Erstellung eines Golden Data Sets durch strukturierte und standardisierte Datenerfassung.	Überprüfung und Weiterentwicklung regulatorischer und ethischer Leitlinien für den Einsatz von KI-gestützter Datenerfassung. Wissenschaftliche Begleitung und Bewertung des Golden Data Sets hinsichtlich Qualität und Datenschutz. Austausch mit GLAUX GROUP und Institutionen zur Definition praxisgerechter Anforderungen an KI-Modelle.

2027	Inbetriebnahme der KI-gestützten Datenerfassung im Modul <i>Journal</i> . Modelltraining und Optimierung.	Pilotierung der KI-gestützten Datenerfassung im Modul <i>Journal</i> und kontinuierliche Rückmeldung zur Ergebnisqualität.	Bewertung der ersten Praxisergebnisse zur KI-gestützten Datenerfassung und Berichterstellung aus ethischer, datenschutzrechtlicher und regulatorischer Perspektive. Entwicklung von Qualitätskriterien für KI-generierte Berichte.
	Extraktion und Validierung von Daten aus Journaleinträgen. Erstellung eines PoC für die KI-gestützte Berichterstellung und Durchführung erster Tests.	Durchführung von Tests zur Praxistauglichkeit der automatisierten Datenverarbeitung und Validierung.	Wissenschaftliche Begleitung der Pilotierung und erste empirische Studien zur Praxistauglichkeit der KI-gestützten Berichterstellung.
	Konzeption weiterer KI-Funktionen.	Mitwirkung bei der Entwicklung und Bewertung des PoC für die KI-gestützte Berichterstellung.	Wissenschaftliche Analyse der Auswirkungen auf die Fachpraxis in der Sozialen Arbeit.
		Identifikation weiterer Einsatzmöglichkeiten und praxisrelevanter Anforderungen für zukünftige KI-Funktionen.	
2028	Umsetzung und Inbetriebnahme umfassender KI-Funktionen in <i>socialweb</i> .	Integration der KI-gestützten Funktionen in den Arbeitsalltag und Anpassung interner Prozesse.	Überprüfung der regulatorischen Konformität und ethischen Auswirkungen.
		Evaluierung der Effizienz und Qualität der KI-gestützten Berichterstellung.	Evaluierung der gesellschaftlichen Akzeptanz und Wahrnehmung der KI-gestützten Berichterstellung.
		Schulung der Mitarbeitenden im Umgang mit den neuen KI-gestützten Tools.	Ableitung von Handlungsempfehlungen für Institutionen und politische Entscheidungsträger basierend auf den Evaluierungsergebnissen.
		Bereitstellung von Praxisfeedback zur kontinuierlichen Verbesserung der KI-Funktionen.	
2029+	Langfristige Weiterentwicklung und Optimierung der KI-Systeme. Aktive Mitgestaltung im Austausch mit allen Akteurinnen und Akteuren.	Kontinuierliche Verbesserung der internen Prozesse zur Nutzung der KI-gestützten Berichterstellung.	Langfristige Begleitung der KI-Entwicklung aus regulatorischer, ethischer und gesellschaftlicher Perspektive.
		Weiterentwicklung der Dokumentationsstandards basierend auf den Erfahrungen mit KI.	Regelmässige Überprüfung der rechtlichen Rahmenbedingungen und Anpassung an neue technologische Entwicklungen.
		Regelmässige Evaluation der KI-Systeme hinsichtlich Qualität, Effizienz und ethischer Aspekte.	Förderung interdisziplinärer Forschung zu den Auswirkungen von KI in der Sozialen Arbeit.
		Aktive Mitgestaltung neuer KI-Funktionen durch den Austausch mit allen Akteurinnen und Akteuren.	

Tabelle 32: Roadmap zur Implementierung von KI-gestützten Funktionen in *socialweb* (2025–2029+).

Die Roadmap zeigt, dass die erfolgreiche Implementierung von KI in der Sozialen Arbeit nicht nur eine technologische Herausforderung darstellt, sondern auch organisatorische, ethische und regulatorische Fragen aufwirft. Neben der Weiterentwicklung der KI-Modelle ist insbesondere die Sicherstellung einer hohen Datenqualität ein zentraler Erfolgsfaktor. Dies führt zu weiteren zentralen Themen, die in den kommenden Jahren vertieft behandelt werden müssen.

Ein essenzieller Erfolgsfaktor für die Weiterentwicklung KI-gestützter Berichterstellung bleibt die Optimierung der Datenqualität. Klare Dokumentationsstandards und technische Unterstützung bei der Datenerfassung könnten die Grundlage für eine verbesserte KI-Nutzung schaffen. Zudem sollte die Entwicklung fachspezifischer KI-Modelle weiter untersucht werden. Derzeit sind viele bestehende

Modelle nicht optimal auf die Anforderungen der Sozialen Arbeit zugeschnitten – insbesondere in der mehrsprachigen Schweiz besteht hier noch Forschungs- und Entwicklungsbedarf.

Die Diskussion über den verantwortungsvollen Einsatz von KI in der Sozialen Arbeit steht erst am Anfang und erfordert eine kontinuierliche Reflexion und Anpassung an neue technologische und gesellschaftliche Entwicklungen. Langfristig könnte KI nicht nur bei der Berichterstellung, sondern auch in anderen Bereichen der Sozialen Arbeit unterstützend eingesetzt werden – beispielsweise in der Fallanalyse, Mustererkennung oder administrativen Entlastung. Dies setzt jedoch voraus, dass Fachpersonen aktiv in den Entwicklungsprozess eingebunden werden und die Technologie gezielt als Assistenzsystem und nicht als Ersatz für professionelle Entscheidungen genutzt wird.

Die kommenden Jahre werden zeigen, inwiefern KI in der Sozialen Arbeit nachhaltig etabliert werden kann. Entscheidend wird sein, dass technologische Entwicklungen mit den fachlichen, ethischen und sozialen Anforderungen in Einklang gebracht werden. Eine enge Zusammenarbeit zwischen Praxis, Forschung und Technologieunternehmen wird dabei zentral sein.

12 Tabellenverzeichnis

Tabelle 1: Architektur-Varianten von Transformatoren [18].	21
Tabelle 2: Methoden für optimales Prompting [30, Kap. Overview of the Five Principles of Prompting ff.], [16, Kap. The Potential Complexity of a Prompt].	29
Tabelle 3: Dimensionen für die Bewertung von Datenqualität [33, Kap. Assessing data quality].	32
Tabelle 4: Themen und relevante Kapitel für die Beantwortung der 1. technischen Forschungsfrage.	37
Tabelle 5: Themen und relevante Kapitel für die Beantwortung der 2. technischen Forschungsfrage.	38
Tabelle 6: Themen und relevante Kapitel für die Beantwortung der 3. technischen Forschungsfrage.	38
Tabelle 7: Themen und relevante Kapitel für die Beantwortung der 4. technischen Forschungsfrage.	39
Tabelle 8: Themen und relevante Kapitel für die Beantwortung der 5. technischen Forschungsfrage.	39
Tabelle 9: Themen und relevante Kapitel für die Beantwortung der 1. Forschungsfrage im Zusammenhang mit der Sozialen Arbeit.	44
Tabelle 10: Themen und relevante Kapitel für die Beantwortung der 2. Forschungsfrage im Zusammenhang mit der Sozialen Arbeit.	45
Tabelle 11: Themen und relevante Kapitel für die Beantwortung der 3. Forschungsfrage im Zusammenhang mit der Sozialen Arbeit.	45
Tabelle 12: Bewertungskriterien der Stossrichtung und Gewichtungen.	47
Tabelle 13: Bewertung der Stossrichtung 1.	51
Tabelle 14: Bewertung Stossrichtung 3.	53
Tabelle 15: Quantitative Bewertung der Daten.	55
Tabelle 16: Qualitative Bewertung der Daten nach Jones und DAMA.	57
Tabelle 17: Herausforderungen im MLOps-Prozess nach Farah im Bereich <i>Data</i> .	66
Tabelle 18: Herausforderungen im MLOps-Prozess nach Farah im Bereich <i>ML</i> .	66
Tabelle 19: Aktuelle Felder und Funktionen beim Erstellen von Journaleinträgen in <i>socialweb</i> .	69
Tabelle 20: Validierungsoptionen beim Speichern von Journaleinträgen in <i>socialweb</i> .	70
Tabelle 21: Aufgaben und Verantwortlichkeiten der Akteurinnen und Akteure entlang des MLOps-Prozesses.	74
Tabelle 22: Themen und relevante Kapitel für die Beantwortung der Hauptforschungsfrage.	80
Tabelle 23: Themen und relevante Kapitel für die Beantwortung der 1. technischen Frage.	80
Tabelle 24: Themen und relevante Kapitel für die Beantwortung der 2. technischen Frage.	81
Tabelle 25: Themen und relevante Kapitel für die Beantwortung der 3. technischen Frage.	81
Tabelle 26: Themen und relevante Kapitel für die Beantwortung der 4. technischen Frage.	81
Tabelle 27: Themen und relevante Kapitel für die Beantwortung der 5. technischen Frage.	82
Tabelle 28: Themen und relevante Kapitel für die Beantwortung der 1. Frage im Zusammenhang mit der Sozialen Arbeit.	83
Tabelle 29: Themen und relevante Kapitel für die Beantwortung der 2. Frage im Zusammenhang mit der Sozialen Arbeit.	83
Tabelle 30: Themen und relevante Kapitel für die Beantwortung der 3. Frage im Zusammenhang mit der Sozialen Arbeit.	84
Tabelle 31: Themen und relevante Kapitel für die Beantwortung der wirtschaftlichen Fragen.	84
Tabelle 32: Roadmap zur Implementierung von KI-gestützten Funktionen in <i>socialweb</i> (2025–2029+).	87

13 Abbildungsverzeichnis

Abbildung 1: Automatisierte Berichterstellung mit KI. Eigene Abbildung, erstellt mit ChatGPT-4o (OpenAI).	1
Abbildung 2: Aufbau der Masterarbeit. Eigene Abbildung.	12
Abbildung 3: KI-Meilensteine in der Entwicklung [12, S. 23]. Eigene Abbildung.	13
Abbildung 4: Hype Cycle für aufkommende KI-Technologien nach Gartner [14].	14
Abbildung 5: Menschliches Leistungsniveau, das durch Technologie erreicht werden kann [15].	15
Abbildung 6: Aktuelle Teildisziplinen der Künstlichen Intelligenz [1, S. 9].	15
Abbildung 7: Erstellung eines Vokabulars [16, Kap. Representing Language as a Bag-of-Words].	16
Abbildung 8: Funktionsweise des <i>Bag-of-Word-Modells</i> [16, Kap. Representing Language as a Bag-of-Words].	17
Abbildung 9: Funktionsweise eines neuronalen Netzwerks [16, Kap. Better Representations with Dense Vector Embeddings].	18
Abbildung 10: Wahrscheinlichkeiten für Wortmerkmale [16, Kap. Better Representations with Dense Vector Embeddings].	18
Abbildung 11: Visuelle Darstellung von Wortbeziehungen im zweidimensionalen Raum [16, Kap. Better Representations with Dense Vector Embeddings].	19
Abbildung 12: Daten führen zu unterschiedlichen Beziehungen im Modell [16, Kap. Types of Embeddings].	19
Abbildung 13: Autoregressive Abfragestruktur eines RNNs [16, Kap. Encoding and Decoding Context with Attention].	20
Abbildung 14: Anzahl Parameter bei GPT-Versionen [16, Kap. Generative Models: Decoder-Only Models].	22
Abbildung 15: Grundschemata für das Modelltraining [16, Kap. Language Models].	22
Abbildung 16: Optimierung des Modells BERT für spezifische Anwendungsfälle [16, Kap. Representation Models: Encoder-Only Models].	23
Abbildung 17: Nutzung eines Golden Data Sets, um nicht gelabelte Daten automatisch zu annotieren [16, Kap. Augmented SBERT].	23
Abbildung 18: Funktionsweise von RAG [16, Kap. From Search to RAG].	24
Abbildung 19: Zusammenfassung der Merkmale beliebter Sprachmodelle [20].	26
Abbildung 20: Entwicklung der MMLU-Leistung bedeutender Sprachmodelle (2022–2024) [23].	26
Abbildung 21: Veraltete Trainingsdaten können zu falschen Antworten führen [24].	27
Abbildung 22: Das LLM wird mittels Prompt aufgefordert, einen Text zu klassifizieren [16, Kap. The Basic Ingredients of a Prompt].	28
Abbildung 23: Anwendung von mehreren Iterationen beim Prompting [16, Kap. The Basic Ingredients of a Prompt].	30
Abbildung 24: Mehrdeutige Prompts liefern unterschiedliche Resultate [31, S. 12].	30
Abbildung 25: Verwendete Metrik-Familien für unterschiedliche NLP-Aufgaben im Jahr 2023 [37, S. 7].	33
Abbildung 26: Anbindung von Closed Source LLMs über eine API [16, Kap. Proprietary, Private Models].	35
Abbildung 27: Implementierung von Open Source LLMs auf der eigenen Umgebung [16, Kap. Open Models].	36
Abbildung 28: Inhaltsanalytische Kommunikationsmodell nach Mayring [48, S. 59].	43
Abbildung 29: Methodik der Masterarbeit. Eigene Abbildung.	46
Abbildung 30: Methodik der PoCs. Eigene Abbildung.	47
Abbildung 31: Abgabe von <i>Valium</i> und <i>Haldol</i> im Jahr 2024. Eigene Abbildung.	59

Abbildung 32: Fehlerhafter Verlauf der Medikamenteneinnahmen im Jahr 2024 [63].....	59
Abbildung 33: KI-Visualisierung « <i>What companies think A.I. looks like</i> » nach Scherpenberg [71].	62
Abbildung 34: KI-Visualisierung « <i>What it actually is</i> » nach Scherpenberg [71].	62
Abbildung 35: MLOps nach Farah [72].	63
Abbildung 36: MLOps nach Farah [72].	65
Abbildung 37: Beteiligte für die Optimierung der Datenbasis. Eigene Darstellung.	67
Abbildung 38: Printscreen <i>Journaleintrag erstellen</i> in <i>socialweb</i>	68
Abbildung 39: Printscreen <i>PoC Automatisierte Arbeitsplanung</i> in <i>socialweb</i>	78

14 Glossar

Adversarial Attack

Technik, bei der gezielt manipulierte Eingaben verwendet werden, um KI-Modelle zu täuschen und unerwartete oder fehlerhafte Ergebnisse zu erzeugen.

Agenten

Software oder Systeme, die eigenständig Aufgaben ausführen und mit ihrer Umgebung interagieren, oft im Kontext von KI oder Multi-Agenten-Systemen.

AI (-> KI, Künstliche Intelligenz)

Annotation

Manuelle oder automatisierte Kennzeichnung von Daten, um sie für das Training von KI-Modellen nutzbar zu machen, z. B. durch Kategorisierung oder Markierung relevanter Informationen.

Applikation

Softwarelösung oder Anwendung, die spezifische Funktionen bereitstellt, beispielsweise im Bereich der KI-gestützten Berichterstellung.

Aufmerksamkeit (Attention Mechanism)

Technik in neuronalen Netzwerken, die es Modellen ermöglicht, sich selektiv auf relevante Teile der Eingabe zu konzentrieren, um bessere Vorhersagen zu treffen.

Autoregressive Modelle

KI-Modelle, die den nächsten Wert in einer Sequenz auf Basis vorheriger Werte vorhersagen, häufig in Sprachmodellen wie GPT verwendet.

Benchmark

Standardisierte Tests oder Vergleichswerte zur Bewertung der Leistung von KI-Modellen oder Softwarelösungen.

Bias (Verzerrung)

Systematische Fehler oder Ungleichgewichte in Daten oder Modellen, die zu unfairen oder ungenauen Ergebnissen führen können.

Black-Box-Modell

KI-System, dessen Entscheidungsprozesse für Fachpersonen aus der IT und der Sozialen Arbeit nicht vollständig nachvollziehbar sind.

Chatbot

Software, die natürliche Sprache verarbeitet und automatisierte Gespräche mit Nutzenden führt, oft basierend auf KI-Modellen.

ChatGPT

Ein generatives Sprachmodell von OpenAI, das auf der GPT-Architektur basiert und Texteingaben in natürlicher Sprache verarbeitet und beantwortet.

CISO (Chief Information Security Officer)

Führungskraft, die für die Informationssicherheit einer Organisation verantwortlich ist.

Dataset (Datensatz)

Sammlung von strukturierten oder unstrukturierten Daten, die für das Training oder die Evaluierung von KI-Modellen genutzt werden.

Datenethik

Disziplin, die sich mit moralischen und gesellschaftlichen Aspekten der Datennutzung und KI-Anwendungen befasst.

Datenqualität

Maß für die Verlässlichkeit, Konsistenz und Vollständigkeit von Daten, das entscheidend für die Leistungsfähigkeit von KI-Modellen ist.

Datenschutz

Rechtliche und technische Massnahmen zum Schutz personenbezogener Daten vor unbefugtem Zugriff oder Missbrauch.

Datensouveränität

Kontrolle von Individuen oder Organisationen über ihre eigenen Daten und deren Nutzung.

Decoder

Teil eines neuronalen Netzwerks, das kodierte Daten in eine verständliche oder verwertbare Form überführt, z. B. in Übersetzungsmodellen.

Dossier (Akte)

Sammlung von Falldokumentationen, Berichten und weiteren Informationen zu einer betreuten Person in der Sozialen Arbeit.

Embeddings

Numerische Repräsentationen von Wörtern oder Konzepten in einem mehrdimensionalen Raum, die semantische Ähnlichkeiten erfassen.

Encoder

Neuronale Netzwerkkomponente, die Eingaben verarbeitet und in eine kompakte Repräsentation umwandelt.

Entitäten

Eindeutige Objekte oder Begriffe, die in Texten identifiziert und klassifiziert werden, z. B. Namen oder Orte.

Halluzinationen (KI)

Falsche oder erfundene Informationen, die ein KI-Modell generiert, obwohl sie nicht aus den Trainingsdaten ableitbar sind.

Interdisziplinär

Zusammenarbeit zwischen verschiedenen Fachbereichen, z. B. zwischen Sozialer Arbeit, Informatik und Ethik in der KI-Entwicklung.

Iterativ

Vorgehensweise, bei der Prozesse schrittweise verbessert und wiederholt werden, um optimale Ergebnisse zu erzielen.

KI (Künstliche Intelligenz)

Technologie, die Maschinen befähigt, menschenähnliche Entscheidungsprozesse oder Mustererkennung durch Algorithmen und Datenverarbeitung nachzuahmen.

Kindeswohlgefährdung

Situation, in der das physische oder psychische Wohl eines Kindes ernsthaft bedroht ist.

Korpora (Singular: Korpus)

Sammlung von Texten oder Sprachdaten, die für das Training und die Analyse von KI-Modellen verwendet wird.

LLM (Large Language Model)

Grosse Sprachmodelle, die auf umfangreichen Textkorpora trainiert wurden und komplexe Sprachverarbeitungsaufgaben lösen können.

MLOps (Machine Learning Operations)

Methoden und Best Practices zur effizienten Entwicklung, Implementierung und Wartung von KI-gestützten Modellen in Unternehmen.

Modelltraining

Prozess, in dem ein KI-Modell durch Mustererkennung in Daten trainiert wird, um Vorhersagen oder Entscheidungen zu verbessern.

NLP (Natural Language Processing)

Technologie zur automatisierten Verarbeitung und Analyse natürlicher Sprache.

Persona

Fiktive Nutzerprofile, die typische Eigenschaften und Bedürfnisse einer Zielgruppe beschreiben und in der KI-Entwicklung verwendet werden.

Proof-of-Concept (PoC)

Experiment oder Testphase, um die Machbarkeit einer Idee oder Technologie zu überprüfen.

Prompting

Technik zur Steuerung von KI-Modellen durch gezielte Texteingaben, um spezifische Antworten zu erhalten.

Prototyp

Vorläufige Version einer Software oder KI-Anwendung zur Evaluierung und Weiterentwicklung.

Retrieval-Augmented Generation (RAG)

Verfahren, das Dokumentensuche mit generativer KI kombiniert, um präzisere und kontextbezogene Antworten zu erzeugen.

Semantik

Bedeutung von Wörtern und Sätzen, die für die Sprachverarbeitung in KI-Systemen analysiert wird.

Sentimentanalyse

Methode zur Bestimmung der emotionalen Tonalität eines Textes, z. B. zur Analyse von Kundenfeedback.

Socialweb

Softwarelösung, die in der Sozialen Arbeit zur Dokumentation und Fallverwaltung genutzt wird.

Solver

Mathematische Algorithmen, die Optimierungsprobleme lösen, z. B. für Schichtplanung oder Entscheidungsfindung.

Sprachmodell

KI-Modell, das darauf trainiert ist, natürliche Sprache zu verstehen und zu generieren.

Token

Einzelne Elemente eines Textes (z. B. Wörter oder Zeichen), die in der KI-gestützten Sprachverarbeitung analysiert werden.

Transformatoren, Transformer-Modelle

Neuronale Netzwerkarchitektur, die auf Selbst-aufmerksamkeit basiert und für Aufgaben wie maschinelle Übersetzung oder Textgenerierung optimiert ist.

Varianz

Statistisches Mass für die Streuung oder Variabilität von Daten, das in der KI-Analyse relevant ist.

Verzerrung (-> Bias)**Wordcloud**

Visuelle Darstellung von Wortfrequenzen in einem Textkorpus, bei der häufig vorkommende Begriffe grösser dargestellt werden.

15 Literaturverzeichnis

- [1] H. von der Gracht, A. Huber-Straßer, M. Schüller, N. Müller, P. Lichtenau, und H. Zühlke, *Wertschöpfung neu gedacht: Von Humanoiden, KI's und Kollege Roboter*. 2018.
- [2] A. Gillhuber, G. Kauermann, und W. Hauner, Hrsg., *Künstliche Intelligenz und Data Science in Theorie und Praxis: Von Algorithmen und Methoden zur praktischen Umsetzung in Unternehmen*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2023. doi: 10.1007/978-3-662-66278-6.
- [3] R. Altenburger und R. Schmidpeter, Hrsg., *CSR und Künstliche Intelligenz*. in Management-Reihe Corporate Social Responsibility. Berlin, Heidelberg: Springer Berlin Heidelberg, 2021. doi: 10.1007/978-3-662-63223-9.
- [4] UNESCO, «Recommendation on the Ethics of Artificial Intelligence». Zugegriffen: 1. November 2024. [Online]. Verfügbar unter: <https://unesdoc.unesco.org/ark:/48223/pf0000381137>
- [5] R. Chatila und J. C. Havens, «The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems», in *Robotics and Well-Being*, Bd. 95, M. I. Aldinhas Ferreira, J. Silva Sequeira, G. Singh Virk, M. O. Tokhi, und E. E. Kadar, Hrsg., in Intelligent Systems, Control and Automation: Science and Engineering, vol. 95. , Cham: Springer International Publishing, 2019, S. 11–16. doi: 10.1007/978-3-030-12524-0_2.
- [6] G. A. Linnemann, J. Löhe, und B. Rottkemper, «Bedeutung von Künstlicher Intelligenz in der Sozialen Arbeit», doi: <https://doi.org/10.1007/s12592-023-00455-7>.
- [7] «Data-Literacy-Charta | Stifterverband». Zugegriffen: 31. Oktober 2024. [Online]. Verfügbar unter: <https://www.stifterverband.org/charta-data-literacy>
- [8] *Verordnung (EU) 2024/1689 des Europäischen Parlaments und des Rates vom 13. Juni 2024 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz und zur Änderung der Verordnungen (EG) Nr. 300/2008, (EU) Nr. 167/2013, (EU) Nr. 168/2013, (EU) 2018/858, (EU) 2018/1139 und (EU) 2019/2144 sowie der Richtlinien 2014/90/EU, (EU) 2016/797 und (EU) 2020/1828 (Verordnung über künstliche Intelligenz) (Text von Bedeutung für den EWR)*. 2024. Zugegriffen: 25. Oktober 2024. [Online]. Verfügbar unter: <http://data.europa.eu/eli/reg/2024/1689/oj/deu>
- [9] «Bundesgesetz über den Datenschutz (Datenschutzgesetz, DSG)», 2020.
- [10] «ISO/IEC 27001», *Wikipedia*. 24. Juli 2024. Zugegriffen: 31. Oktober 2024. [Online]. Verfügbar unter: https://de.wikipedia.org/w/index.php?title=ISO/IEC_27001&oldid=247030716
- [11] J. C. Costa, T. Roxo, H. Proença, und P. R. M. Inácio, «How Deep Learning Sees the World: A Survey on Adversarial Attacks & Defenses», 18. Mai 2023, *arXiv*: arXiv:2305.10862. doi: 10.48550/arXiv.2305.10862.
- [12] W. Ertel, *Grundkurs Künstliche Intelligenz: Eine praxisorientierte Einführung*. in Computational Intelligence. Wiesbaden: Springer Fachmedien Wiesbaden, 2021. doi: 10.1007/978-3-658-32075-1.
- [13] M. Heinlein und N. Huchler, Hrsg., *Künstliche Intelligenz, Mensch und Gesellschaft: Soziale Dynamiken und gesellschaftliche Folgen einer technologischen Innovation*. Wiesbaden: Springer Fachmedien Wiesbaden, 2024. doi: 10.1007/978-3-658-43521-9.
- [14] «Gartner Places Generative AI on the Peak of Inflated Expectations on the 2023 Hype Cycle for Emerging Technologies», Gartner. Zugegriffen: 25. Oktober 2024. [Online]. Verfügbar unter: <https://www.gartner.com/en/newsroom/press-releases/2023-08-16-gartner-places-generative-ai-on-the-peak-of-inflated-expectations-on-the-2023-hype-cycle-for-emerging-technologies>
- [15] M. Chui u. a., «2023The economic potential of generative AI: The next productivity frontier», McKinsey & Company, Juni 2023.
- [16] J. Alammam und M. Grootendorst, *Hands-On Large Language Models*. Zugegriffen: 30. Januar 2025. [Online]. Verfügbar unter: <https://learning.oreilly.com/library/view/hands-on-large-language/9781098150952/>
- [17] A. Vaswani u. a., «Attention Is All You Need», 2. August 2023, *arXiv*: arXiv:1706.03762. doi: 10.48550/arXiv.1706.03762.
- [18] M. Hoque, «A Comprehensive Overview of Transformer-Based Models: Encoders, Decoders, and More», Medium. Zugegriffen: 29. November 2024. [Online]. Verfügbar unter: <https://medium.com/@minh.hoque/a-comprehensive-overview-of-transformer-based-models-encoders-decoders-and-more-e9bc0644a4e5>

- [19] «Common Crawl», *Wikipedia*. 23. November 2024. Zugegriffen: 30. November 2024. [Online]. Verfügbar unter: https://en.wikipedia.org/w/index.php?title=Common_Crawl&oldid=1259054634
- [20] A. Halevy, P. Norvig, und N. Fernando, «The Unreasonable Effectiveness of Data», *Intell. Syst. IEEE*, Bd. 24, S. 8–12, Mai 2009, doi: 10.1109/MIS.2009.36.
- [21] «Hugging Face – The AI community building the future.» Zugegriffen: 30. November 2024. [Online]. Verfügbar unter: <https://huggingface.co/>
- [22] D. J. Lipenkova, «Choosing the right language model for your NLP use case», Medium. Zugegriffen: 21. Oktober 2024. [Online]. Verfügbar unter: <https://towardsdatascience.com/choosing-the-right-language-model-for-your-nlp-use-case-1288ef3c4929>
- [23] «Major Large Language Models (LLMs)». Zugegriffen: 27. November 2024. [Online]. Verfügbar unter: <https://informationisbeautiful.net/visualizations/the-rise-of-generative-ai-large-language-models-llms-like-chatgpt/>
- [24] «Chatbot Arena (formerly LMSYS): Free AI Chat to Compare & Test Best AI Chatbots». Zugegriffen: 28. November 2024. [Online]. Verfügbar unter: <https://lmarena.ai/>
- [25] «Introducing ChatGPT search». Zugegriffen: 30. November 2024. [Online]. Verfügbar unter: <https://openai.com/index/introducing-chatgpt-search/>
- [26] C. Rigotti, A. Puttick, E. Fosch-Villaronga, und M. Kurpicz-Briki, «Mitigating Diversity Biases of AI in the Labor Market», 2023.
- [27] «What Are AI Hallucinations? | IBM». Zugegriffen: 1. Dezember 2024. [Online]. Verfügbar unter: <https://www.ibm.com/topics/ai-hallucinations>
- [28] N. Dziri, S. Milton, M. Yu, O. Zaiane, und S. Reddy, «On the Origin of Hallucinations in Conversational Models: Is it the Datasets or the Models?», 17. April 2022, *arXiv*: arXiv:2204.07931. doi: 10.48550/arXiv.2204.07931.
- [29] Z. Muhammad Zayyanu Ph.D., «Revolutionising Translation Technology: A Comparative Study of Variant Transformer Models - BERT, GPT, and T5», *Comput. Sci. Eng. Int. J.*, Bd. 14, S. 15–27, Juni 2024, doi: 10.5121/cseij.2024.14302.
- [30] *Prompt Engineering for Generative AI*. Zugegriffen: 25. Oktober 2024. [Online]. Verfügbar unter: <https://learning.oreilly.com/library/view/prompt-engineering-for/9781098153427/>
- [31] I. Stoica u. a., «Specifications: The missing link to making the development of LLM systems an engineering discipline», 16. Dezember 2024, *arXiv*: arXiv:2412.05299. doi: 10.48550/arXiv.2412.05299.
- [32] B. Weber, Hrsg., *Data Governance: Nachhaltige Geschäftsmodelle und Technologien im europäischen Rechtsrahmen*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2023. doi: 10.1007/978-3-662-67556-4.
- [33] A. Jones, *Data quality in the Age of AI: building a foundation for AI strategy and data culture*. Place of publication not identified: Packt Publishing, 2024.
- [34] N. Askham u. a., «The six primary dimensions for data quality assessment: Defining data quality dimensions». Zugegriffen: 9. Dezember 2024. [Online]. Verfügbar unter: <https://www.sbctc.edu/resources/documents/collegesstaff/commissions-councils/dgc/data-quality-deminions.pdf>.
- [35] «MasterArbeit/Datenanalyse/sentimentanalyse.py at main · Matthias-Andenmatten/MasterArbeit». Zugegriffen: 17. Februar 2025. [Online]. Verfügbar unter: <https://github.com/Matthias-Andenmatten/MasterArbeit/blob/main/Datenanalyse/sentimentanalyse.py>
- [36] K. P. Gunasekaran, «Exploring Sentiment Analysis Techniques in Natural Language Processing: A Comprehensive Review».
- [37] P. Schmidtova u. a., «Automatic Metrics in Natural Language Generation: A survey of Current Evaluation Practices».
- [38] E. Reiter, «A Structured Review of the Validity of BLEU», *Comput. Linguist.*, Bd. 44, Nr. 3, S. 393–401, Sep. 2018, doi: 10.1162/coli_a_00322.
- [39] C. Van Der Lee, A. Gatt, E. Van Miltenburg, und E. Krahmer, «Human evaluation of automatically generated text: Current trends and best practice guidelines», *Comput. Speech Lang.*, Bd. 67, S. 101151, Mai 2021, doi: 10.1016/j.csl.2020.101151.
- [40] A. Celikyilmaz, A. Bosselut, X. He, und Y. Choi, «Deep Communicating Agents for Abstractive Summarization», in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, New Orleans, Louisiana: Association for Computational Linguistics, 2018, S. 1662–1675. doi: 10.18653/v1/N18-1150.

- [41] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, D. Pedreschi, und F. Giannotti, «A Survey Of Methods For Explaining Black Box Models», 21. Juni 2018, *arXiv*: arXiv:1802.01933. doi: 10.48550/arXiv.1802.01933.
- [42] «ChatGPT Pricing». Zugriffen: 1. Dezember 2024. [Online]. Verfügbar unter: <https://openai.com/chatgpt/pricing/>
- [43] G. Schuh und C. Dölle, *Sustainable Innovation: nachhaltig Werte schaffen*, 2. Auflage. Berlin [Heidelberg]: Springer Vieweg, 2021.
- [44] C. Kaminsky, «Digitale Transformation Sozialer Arbeit? – Ethische Orientierungen auf neuem Terrain», 2021.
- [45] R. Gutwald, J. Burghardt, M. Kraus, M. Reder, R. Lehmann, und N. Müller, «Soziale Konflikte und Digitalisierung», 2021.
- [46] B. Görder, «Die Ethik der Sozialen Arbeit vor professionsbezogenen und gesellschaftlichen Herausforderungen durch ‚künstliche Intelligenz‘», 2021.
- [47] U. Reichmann, *Schreiben und Dokumentieren in der Sozialen Arbeit: Struktur, Orientierung und Reflexion für die berufliche Praxis*. in UTB Soziale Arbeit, no. 4579. Opladen Berlin Toronto: Verlag Barbara Budrich, 2016. doi: 10.36198/9783838545790.
- [48] P. Mayring, *Qualitative Inhaltsanalyse: Grundlagen und Techniken*, 12., Vollständig überarbeitete und aktualisierte Aufl. in Beltz Pädagogik. Weinheim: Beltz, 2015.
- [49] «Künstliche Intelligenz programmieren: Die besten Coding-Sprachen für KI», Computerwoche. Zugriffen: 4. November 2024. [Online]. Verfügbar unter: <https://www.computerwoche.de/article/2795556/die-besten-coding-sprachen-fuer-ki.html>
- [50] powered by sitesystem info@webways.ch <https://www.sitesystem.ch> | created by webways ag, basel, <https://www.webways.ch>, «INSOS - Der Branchenverband der Dienstleister für Menschen mit Behinderung», INSOS - Home. Zugriffen: 8. November 2024. [Online]. Verfügbar unter: <http://www.insos.ch>
- [51] J. H. Choi, K. E. Hickman, A. Monahan, und D. Schwarcz, «ChatGPT Goes to Law School», 23. Januar 2023, *Social Science Research Network, Rochester, NY*: 4335905. doi: 10.2139/ssrn.4335905.
- [52] R. Gozalo-Brizuela und E. C. Garrido-Merchan, «ChatGPT is not all you need. A State of the Art Review of large Generative AI models», 11. Januar 2023, *arXiv*: arXiv:2301.04655. Zugriffen: 8. November 2024. [Online]. Verfügbar unter: <http://arxiv.org/abs/2301.04655>
- [53] L. Tunstall, L. von Werra, und T. Wolf, *Natural Language Processing mit Transformern: Sprachanwendungen mit Hugging Face erstellen*. in Animals. Heidelberg: O'Reilly, 2023.
- [54] «MasterArbeit/Stossrichtung1_Zusammenfassung_alle_Modelle_Resultat at main · Matthias-Andenmatten/MasterArbeit». Zugriffen: 21. November 2024. [Online]. Verfügbar unter: https://github.com/Matthias-Andenmatten/MasterArbeit/blob/main/Evaluation%20der%20Stossrichtung/Stossrichtung1_Zusammenfassung_alle_Modelle_Resultat.txt
- [55] «MasterArbeit/Stossrichtung1_mit_Halluzination_Resultat at main · Matthias-Andenmatten/MasterArbeit». Zugriffen: 21. November 2024. [Online]. Verfügbar unter: https://github.com/Matthias-Andenmatten/MasterArbeit/blob/main/Evaluation%20der%20Stossrichtung/Stossrichtung1_mit_Halluzination_Resultat.txt
- [56] «MasterArbeit/Stossrichtung3_Zusammenfassung_Resultat.txt at main · Matthias-Andenmatten/MasterArbeit». Zugriffen: 21. November 2024. [Online]. Verfügbar unter: https://github.com/Matthias-Andenmatten/MasterArbeit/blob/main/Evaluation%20der%20Stossrichtung/Stossrichtung3_Zusammenfassung_Resultat.txt
- [57] «MasterArbeit/Stossrichtung3_Zusammenfassung_Skript.py at main · Matthias-Andenmatten/MasterArbeit». Zugriffen: 21. November 2024. [Online]. Verfügbar unter: https://github.com/Matthias-Andenmatten/MasterArbeit/blob/main/Evaluation%20der%20Stossrichtung/Stossrichtung3_Zusammenfassung_Skript.py
- [58] «OpenAI Help Center». Zugriffen: 21. November 2024. [Online]. Verfügbar unter: <https://help.openai.com/en/>
- [59] «MasterArbeit/Stossrichtung3_Zusammenfassung_lang_Resultat.txt at main · Matthias-Andenmatten/MasterArbeit». Zugriffen: 21. November 2024. [Online]. Verfügbar unter: https://github.com/Matthias-Andenmatten/MasterArbeit/blob/main/Evaluation%20der%20Stossrichtung/Stossrichtung3_Zusammenfassung_lang_Resultat.txt
- [60] «MasterArbeit/Erstellung von Berichten/K-2021-00003 - Test 1 - Keine Angaben zu Ziel.pdf at main · Matthias-Andenmatten/MasterArbeit». Zugriffen: 24. Januar 2025. [Online]. Verfügbar unter: <https://github.com/Matthias->

- Andenmatten/MasterArbeit/blob/main/Erstellung%20von%20Berichten/K-2021-00003%20-%20Test%201%20-%20Keine%20Angaben%20zu%20Ziel.pdf
- [61] «MasterArbeit/Erstellung von Berichten/K-2021-00003 - Test 3 - Erweiterter Prompt.pdf at main · Matthias-Andenmatten/MasterArbeit». Zugriffen: 24. Januar 2025. [Online]. Verfügbar unter: <https://github.com/Matthias-Andenmatten/MasterArbeit/blob/main/Erstellung%20von%20Berichten/K-2021-00003%20-%20Test%203%20-%20Erweiterter%20Prompt.pdf>
- [62] «MasterArbeit/Erstellung von Berichten/K-2021-00003 - Test 2 - Mehr Informationen und Richtung.pdf at main · Matthias-Andenmatten/MasterArbeit», GitHub. Zugriffen: 24. Januar 2025. [Online]. Verfügbar unter: <https://github.com/Matthias-Andenmatten/MasterArbeit/blob/main/Erstellung%20von%20Berichten/K-2021-00003%20-%20Test%202%20-%20Mehr%20Informationen%20und%20Richtung.pdf>
- [63] «MasterArbeit/Erstellung von Berichten/K-2021-00003 - Test 4 - Ergänzung beim Thema Medikamente.pdf at main · Matthias-Andenmatten/MasterArbeit». Zugriffen: 24. Januar 2025. [Online]. Verfügbar unter: <https://github.com/Matthias-Andenmatten/MasterArbeit/blob/main/Erstellung%20von%20Berichten/K-2021-00003%20-%20Test%204%20-%20Erg%C3%A4nzung%20beim%20Thema%20Medikamente.pdf>
- [64] «MasterArbeit/Erstellung von Berichten/K-2021-00003 - Test 7 - Mehr Kontext und Beispiele zur Verfügung gestellt.pdf at main · Matthias-Andenmatten/MasterArbeit». Zugriffen: 24. Januar 2025. [Online]. Verfügbar unter: <https://github.com/Matthias-Andenmatten/MasterArbeit/blob/main/Erstellung%20von%20Berichten/K-2021-00003%20-%20Test%207%20-%20Mehr%20Kontext%20und%20Beispiele%20zur%20Verf%C3%BCgung%20gestellt.pdf>
- [65] «MasterArbeit/Erstellung von Berichten/K-2021-00079 - Test 12 - Erstellung Koordinationsgespräch mit erweitertem Prompt und Beispiel.pdf at main · Matthias-Andenmatten/MasterArbeit». Zugriffen: 24. Januar 2025. [Online]. Verfügbar unter: <https://github.com/Matthias-Andenmatten/MasterArbeit/blob/main/Erstellung%20von%20Berichten/K-2021-00079%20-%20Test%2012%20-%20Erstellung%20Koordinationsgespr%C3%A4ch%20mit%20erweitertem%20Prompt%20und%20Beispiel.pdf>
- [66] «MasterArbeit/Erstellung von Berichten/K-2021-00060 - Test 10 - Wichtige Bezugspersonen im Kontext.pdf at main · Matthias-Andenmatten/MasterArbeit». Zugriffen: 24. Januar 2025. [Online]. Verfügbar unter: <https://github.com/Matthias-Andenmatten/MasterArbeit/blob/main/Erstellung%20von%20Berichten/K-2021-00060%20-%20Test%2010%20-%20Wichtige%20Bezugspersonen%20im%20Kontext.pdf>
- [67] «MasterArbeit/Erstellung von Berichten/K-2021-00060 - Test 9 - Erstellung Koordinationsgespräch mit erweitertem Prompt und Beispiel.pdf at main · Matthias-Andenmatten/MasterArbeit». Zugriffen: 24. Januar 2025. [Online]. Verfügbar unter: <https://github.com/Matthias-Andenmatten/MasterArbeit/blob/main/Erstellung%20von%20Berichten/K-2021-00060%20-%20Test%209%20-%20Erstellung%20Koordinationsgespr%C3%A4ch%20mit%20erweitertem%20Prompt%20und%20Beispiel.pdf>
- [68] «OpenAI Platform». Zugriffen: 14. Februar 2025. [Online]. Verfügbar unter: <https://platform.openai.com>
- [69] «MasterArbeit/Erstellung von Berichten/K-2021-00003 - Test 8 - Verlauf der Medikamente.pdf at main · Matthias-Andenmatten/MasterArbeit». Zugriffen: 24. Januar 2025. [Online]. Verfügbar unter: <https://github.com/Matthias-Andenmatten/MasterArbeit/blob/main/Erstellung%20von%20Berichten/K-2021-00003%20-%20Test%208%20-%20Verlauf%20der%20Medikamente.pdf>
- [70] D. Lochmatter, «Wie lange wird meine Institution <KI-ignorierend> überleben?», Aug. 2023.
- [71] «Andy Scherpenberg auf LinkedIn: #ai | 324 Kommentare». Zugriffen: 23. Januar 2025. [Online]. Verfügbar unter: https://www.linkedin.com/posts/andyscherpenberg_ai-activity-6863775725061726208-pw8q
- [72] D. Farah, «The Modern MLOps Blueprint», Slalom Data & AI. Zugriffen: 23. Januar 2025. [Online]. Verfügbar unter: <https://medium.com/slalom-data-ai/the-modern-mlops-blueprint-c8322af69d21>

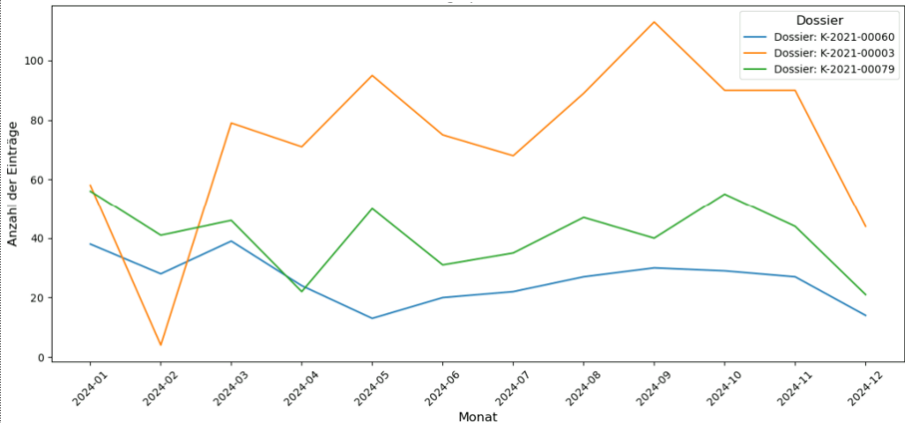
16 Anhang

16.1 Halluzination des Modells «google/pegasus-multi_news»

Prompt: «Fasse das Märchen *Schneewittchen* auf Deutsch zusammen»

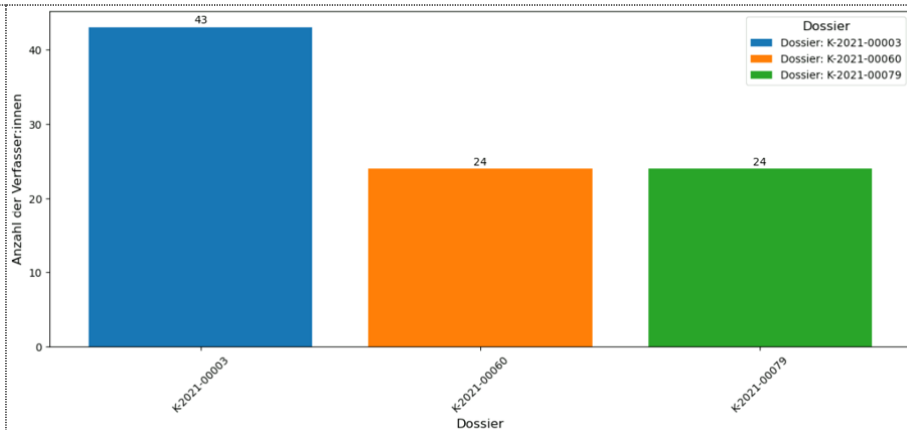
Output: «The latest celebrity to make headlines for his or her Nazi past is none other than Angela Merkel, who was raised in a concentration camp as a child. In an interview with Der Spiegel, the German chancellor revealed that she was raised in a camp called Knigin, where her father was imprisoned and her mother worked as a cook. When her father died, Merkel's mother took over as head of the Knigin camp. "I grew up in a situation where my father was imprisoned and my mother was not," Merkel said. "So I grew up in a situation where my father was imprisoned and my mother was not." Merkel's father was executed in World War II, and her mother was forced to live in a concentration camp for two years before being released. Merkel's grandfather was also imprisoned in the Knigin camp, and her grandfather was executed himself. Merkel's father and grandfather were both members of the Nazi party. In his interview with Der Spiegel, Merkel said that her family was "in a very difficult situation" after her father's death and her grandfather's imprisonment. "I grew up in a situation where my father was imprisoned and my mother was not," she said. "So I grew up in a situation where my father was imprisoned and my grandfather was not. I grew up in a situation where my father was im-
prisoned." »

16.2 Quantitative Datenanalyse

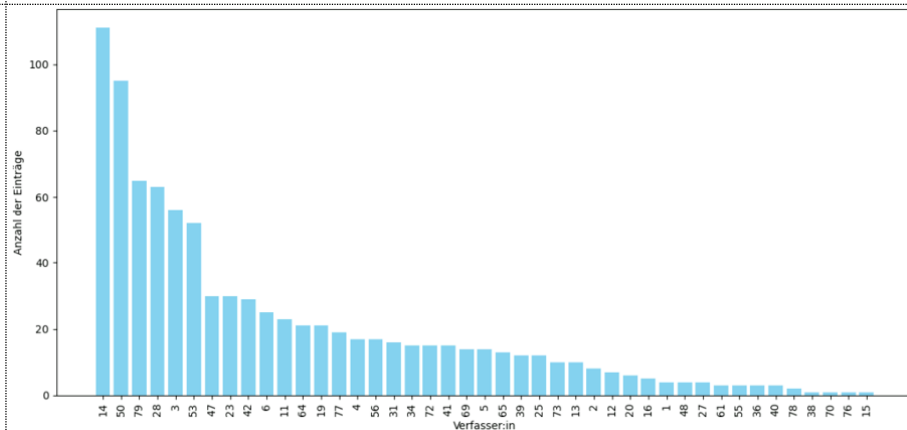
Thema	Beschreibung
Dateityp, Format, Grösse	csv, utf8, 1'958 KB
Spalten und Beschreibung	dossierName der Klient/in entry_atZeitpunkt des Eintrags created_atZeitpunkt der Erstellung bodyInhalt des Eintrags (Text) categoryKategorie, 1:1 markersMarkierung/en 1:n creator_nameName des Verfassers, bzw. der Verfasserin
Anzahl Dossiers	3 (K-2021-00003, K-2021-00060, K-2021-00079)
Anzahl Datensätze	K-2021-00003: 1908 K-2021-00060: 917 K-2021-00079: 1221
Zeitraum	K-2021-00003: 05.02.2022 - 17.12.2024 K-2021-00060: 06.06.2022 - 17.12.2024 K-2021-00079: 30.03.2021 - 16.12.2024
Anzahl Einträge pro Dossier und Monat	

Anzahl Zeichen pro Journal- eintrag und Dossier	Box plot showing the distribution of character counts for three dossiers: K-2021-00003, K-2021-00060, and K-2021-00079. The y-axis represents the number of characters (Anzahl Zeichen) from 0 to 1000. The x-axis represents the dossier number. Each box plot shows the median (green line), interquartile range (blue box), and range (whiskers). Outliers are shown as black dots.
Kategorien	01 Beobachtung, 02 Gesundheit, 03 Freiheitseinschränkende Massnahmen, 04 Vereinbarung, 06 LQ Planung, Anfrage
Anzahl Einträge pro Katego- rie und Dossier	Grouped bar chart showing the number of entries (Anzahl der Einträge) for six categories across three dossiers: K-2021-00003 (blue), K-2021-00060 (orange), and K-2021-00079 (green). The categories are 01 Beobachtung, 02 Gesundheit, 04 Vereinbarung, Anfrage, 06 LQ Planung, and 03 Freiheitseinschränkende Massnahmen. The y-axis ranges from 0 to 400.
Markierungen	1. Autonomie / Selbstbestimmung, 2. Teilhabe, 3. Kommunikation / UK, 4. Familie / Bei- stand, 5. Biografie, 6. Ernährung/Verdauung, 7. Pflege/Medizinisches, 8. Liebe/Sexualität, 9. Herausforderndes Verhalten, 10. Reserve-Medikament (bei Abgabe), 11. Medikamenten- fehler, 12. Relevant für Nachtwache, 13. Strukturierte Beobachtungsphasen, 14. Mobili- tät/Transfer, 15. Lebenspraktische Fähigkeiten, 16. Aktivierung
Anzahl Einträge pro Markie- rung und Dossier	Grouped bar chart showing the number of entries (Anzahl der Einträge) for sixteen markers across three dossiers: K-2021-00003 (blue), K-2021-00060 (orange), and K-2021-00079 (green). The markers are listed on the x-axis. The y-axis ranges from 0 to 400.

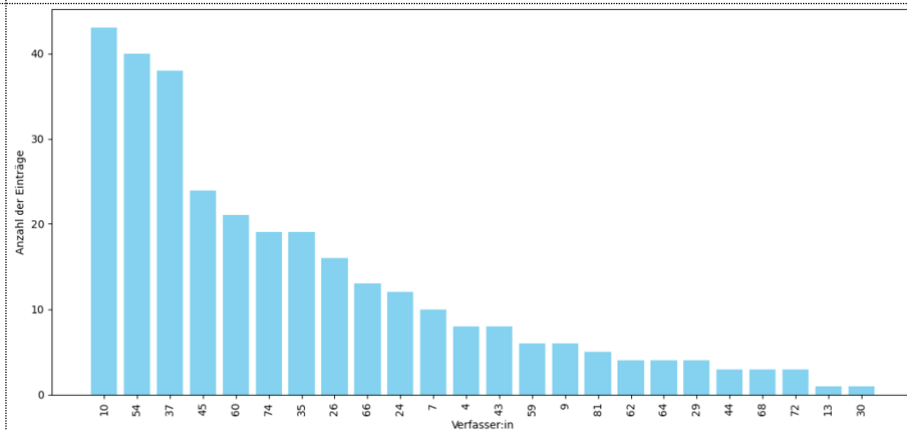
Anzahl Verfasserinnen bzw. Verfasser pro Dossier



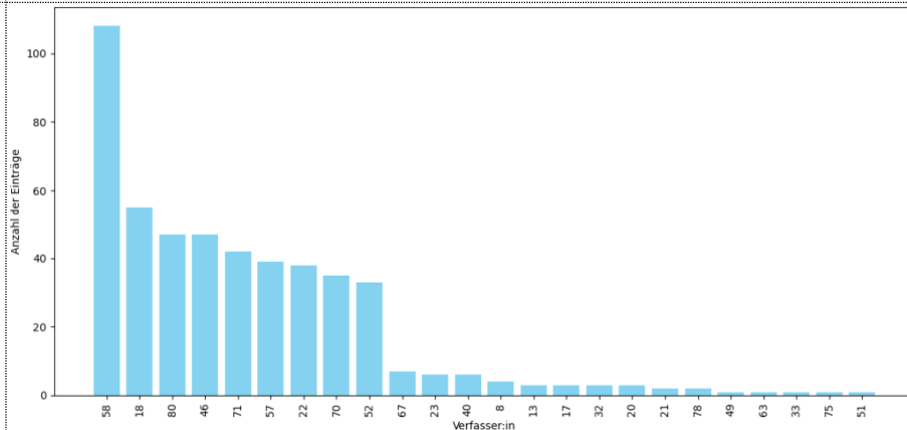
Anzahl Einträge pro Verfasserin bzw. Verfasser und Dossier (K-2021-00003)



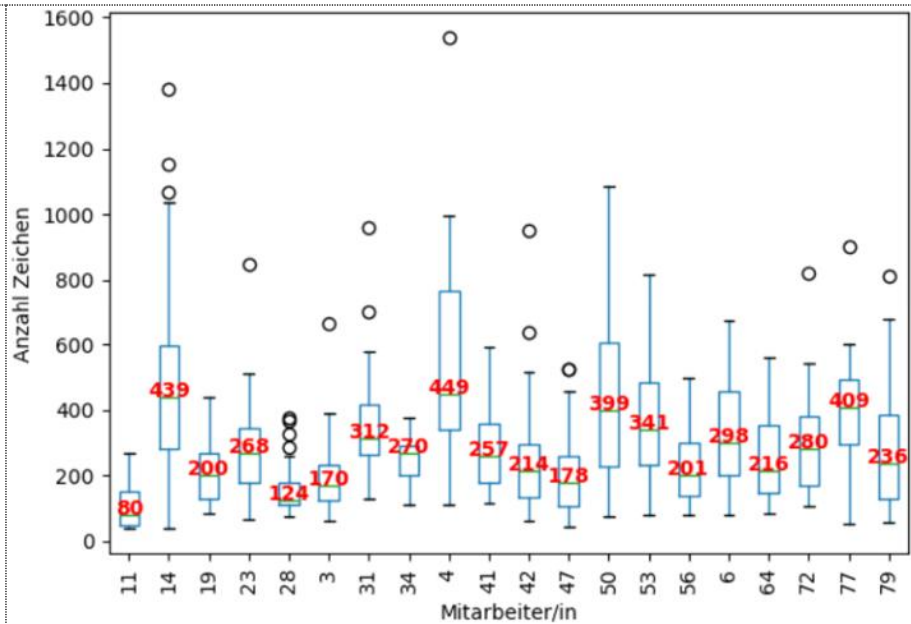
Anzahl Einträge pro Verfasserin bzw. Verfasser und Dossier (K-2021-00060)



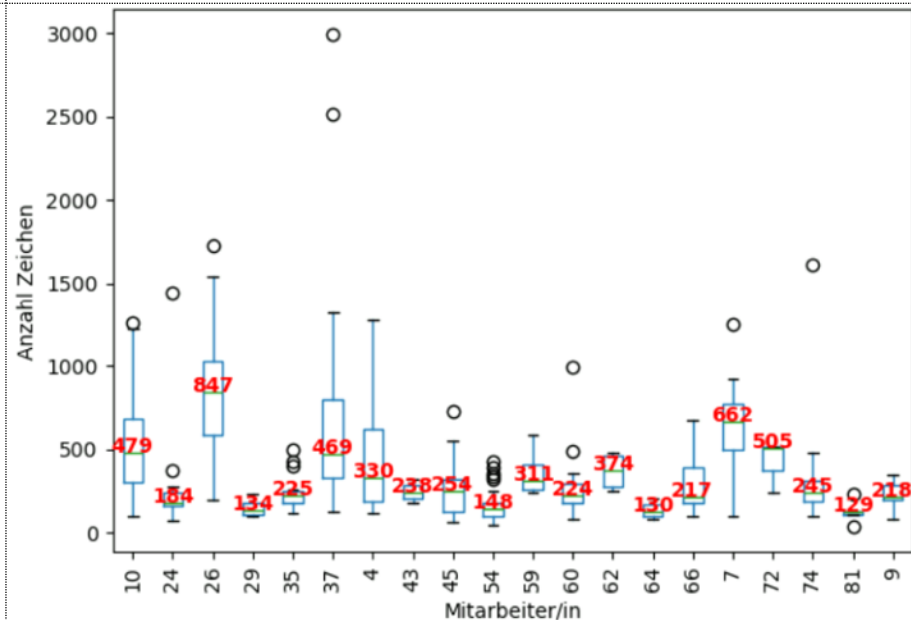
Anzahl Einträge pro Verfasserin bzw. Verfasser und Dossier (K-2021-00079)



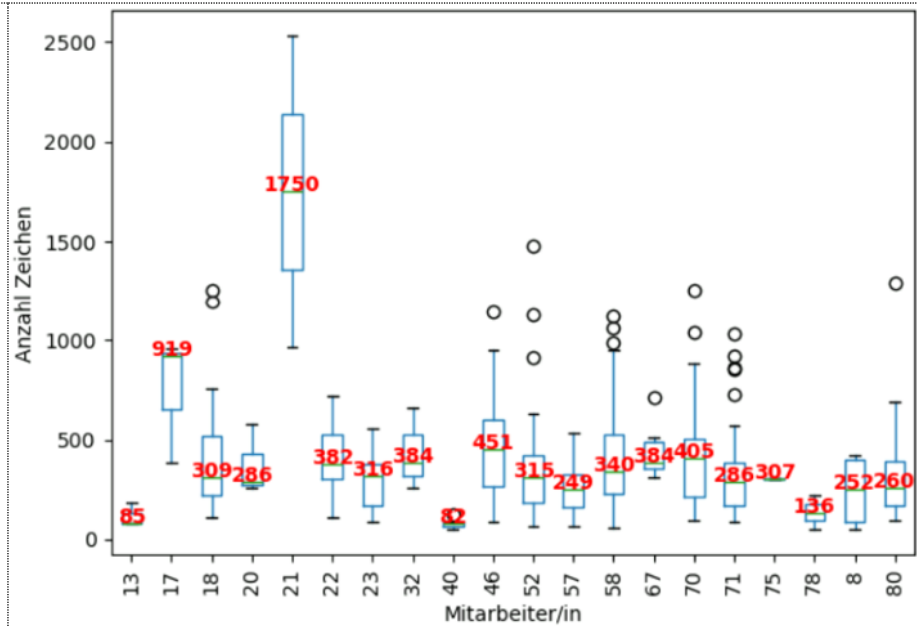
Anzahl Zeichen pro Journal-
eintrag, Dossier (K-2021-
00003) und Mitarbeiterin
bzw. Mitarbeiter (Top 20)



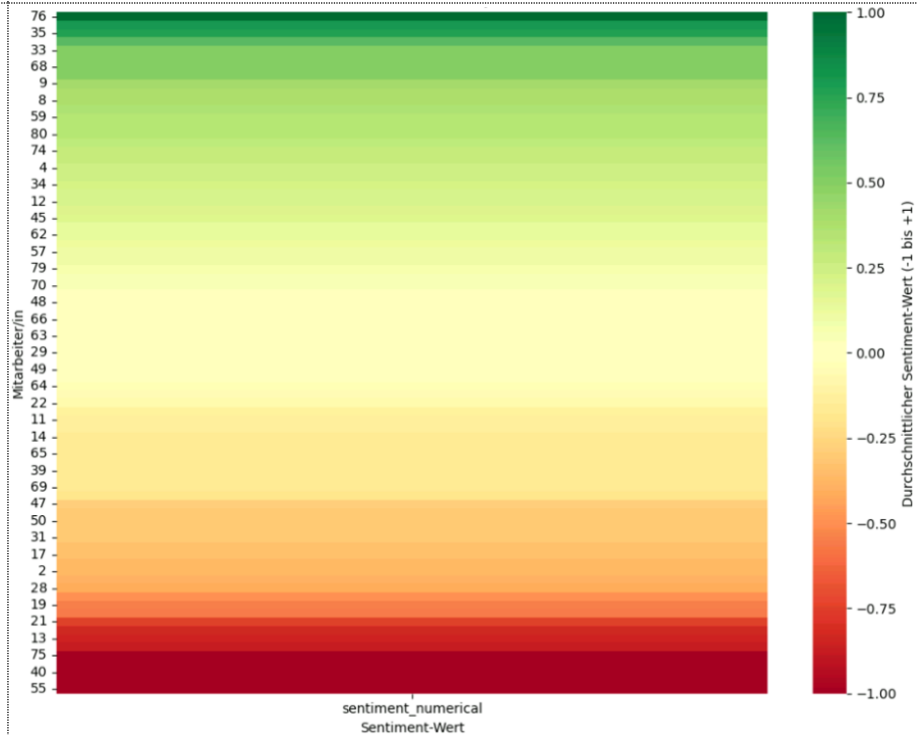
Anzahl Zeichen pro Journal-
eintrag, Dossier (K-2021-
00060) und Mitarbeiterin
bzw. Mitarbeiter (Top 20)



Anzahl Zeichen pro Journal-
eintrag, Dossier (K-2021-
00079) und Mitarbeiterin
bzw. Mitarbeiter (Top 20)



Durchschnittlicher Senti-
ment-Wert pro Mitarbeiterin
bzw. Mitarbeiter



Wordcloud der Top 50 ver-
wendeten Wörter von Dos-
sier (K-2021-00003)



Wordcloud der Top 50 verwendeten Wörter von Dosier (K-2021-00060)

Wordcloud der Top 50 verwendeten Wörter von Dosier (K-2021-00079)

16.3 Qualitative Datenanalyse von Dossier K-2021-00003

Das Dossier K-2021-00003 zeigt eine Vielzahl an Herausforderungen und Unstimmigkeiten in der Dokumentation. Nachfolgend werden die Hauptaspekte strukturiert dargestellt:

Fehlerhafte Schreibweisen und uneinheitliche Sprache

- Vornamen und Nachnamen wurden mehrfach falsch geschrieben, was die Konsistenz und Verlässlichkeit der Daten beeinträchtigten.
- Die Verwendung von Dialekt (z. B. *Zigi*, *Schnitteli*) und Hochdeutsch ist uneinheitlich und kann zu Interpretationsproblemen führen.
- Smileys und private Tonalität (*Wellness Sequenz*) können als unprofessionell aufgefasst werden.

Unklare und uneinheitliche Kategorisierung

- Die Kategorien sind zu grob und bieten keine klare Zuordnung der Einträge. Beispielsweise wird *01 Beobachtung* für thematisch sehr unterschiedliche Inhalte verwendet. Beispielsweise könnten pflegerische Themen wie Duschen, Rasieren oder Kratzspuren in spezifischen Pflegekategorien erfasst werden.
- Medikamente werden im Freitext dokumentiert, häufig ohne klare Dosierungsangaben oder Verweise auf konkrete Massnahmen.

Mängel in der Dokumentationsqualität

- Abkürzungen: Uneinheitliche und unklare Verwendung von Abkürzungen wie *ggts*, *SMFD* oder *BZP* erschweren die Lesbarkeit.
- Rechtschreibung und Grammatik: Die Einträge enthalten Rechtschreib- und Kommafehler, möglicherweise bedingt durch die Mehrsprachigkeit der Mitarbeitenden.

- Privat versus professionell: Manche Einträge wirken privat oder informell, z. B. *Ein Handbad rundete die Wellness-Sequenz ab*.
- Inhalte ohne Zusammenhang: Einige Einträge enthalten Informationen ohne ersichtlichen Bezug zum Bewohnerzustand oder den Massnahmen, z. B. Einkäufe oder Telefonnummern.

Medikationsdokumentation

- Häufig uneinheitliche oder unvollständige Dokumentation von Medikamenten: *Erhält Schmerz gtts. aus der Reserve.* - ohne Dosierung oder konkreten Bezug.
- Abweichende Schreibweisen bei Medikamentennamen oder Einheiten (*gtts* vs. *Tropfen*).

Beobachtungen und Diagnosen

- Diagnosen werden im Freitext anstelle von Beobachtungen formuliert, z. B. *zunehmende Apraxie und Agnosie*
- Beobachtungen wie *tiefe Kratzwunde (?)* oder *Au-au-au* sind unklar formuliert und bedürfen genauerer Spezifizierung.

Teamkommunikation

- Teammitteilungen wie *Bitte abends Haut rückfetten* oder *Wundprotokoll in laminiierter Form ausfüllen* mischen sich mit der Dokumentation und führen zu Unklarheiten.
- Empfehlungen zur Steuerung wie *Mahlzeiten in kleinere Stücke schneiden* werden nicht klar zugewiesen.

Datenschutz

- Die Erwähnung von Namen anderer Klienten, externen Fachpersonen sowie private Telefonnummern im Journal verstösst gegen Datenschutzbestimmungen.

Verbesserungspotenziale

- Standardisierung der Sprache: Einheitlicher Einsatz von Hochdeutsch und Vermeidung von Dialekt oder informellem Stil.
- Abkürzungsverzeichnis: Klare Definition und Konsistenz bei Abkürzungen.
- Klare Kategorien und Module: Trennung zwischen pflegerischen, medizinischen und organisatorischen Themen.
- Schulung der Mitarbeitenden: Verbesserung der Rechtschreibung, der medizinischen Dokumentation und der Einhaltung professioneller Standards.
- Technische Unterstützung: Verwendung von Vorlagen oder Checklisten für Medikamenteneinträge und pflegerische Massnahmen.
- Datenschutzrichtlinien: Sensible Daten wie Telefonnummern und Namen anderer Klienten und Fachpersonen sollten nicht im Journal erfasst werden.

16.4 Qualitative Datenanalyse von Dossier K-2021-00060

Das Dossier K-2021-00060 weist ähnliche Herausforderungen wie K-2021-00003 auf, jedoch mit spezifischen Eigenheiten, die nachfolgend strukturiert dargestellt werden:

Fehlerhafte Schreibweisen und uneinheitliche Sprache

- Der Name des Klienten wird in unterschiedlichen Varianten falsch geschrieben.
- Die Verwendung von Mundart, z. B. *Schätzeli*, beeinträchtigt die Konsistenz und Professionalität der Einträge.
- Einträge enthalten personenbezogene Angaben wie *Auf mich*, *OLA*, die überflüssig sind, wenn OLA die Verfasserin ist.

Datenschutz und externe Inhalte

- Namen von Systemkontakten, anderen Klienten und externen Fachpersonen erscheinen im Journal, was datenschutzrechtlich bedenklich ist.
- Es wird die Verwendung von den Tools *Nicki Dairy* und *Big Point Button* beschrieben. Es müsste geklärt werden, ob diese die Datenschutzbestimmungen erfüllen.

- Der Eintrag vom 20. Februar 2024 enthält die vollständige Adresse und Telefonnummer von einem Lieferanten. Es wird empfohlen die Informationen im Modul *Kontakte* zu erfassen und anschliessend im Modul *Systeme* zu verknüpfen.

Unklare und uneinheitliche Kategorisierung

- Zuordnung von Kategorien: Manche Einträge, z. B. betreffend der Pflege der Bauchfalte (8. Oktober 2024 - 19. Oktober 2024), werden in unterschiedlichen Kategorien erfasst. Entsprechend können die Informationen nicht logisch gefiltert und strukturiert ausgewertet werden.
- Unklare Inhalte in Kategorien: Beispiel vom 28. November 2024 unter *01 Beobachtung* umfasst Themen wie Ziele, Pflege und Therapie - eine klare Trennung wäre sinnvoll.
- Das Journal enthält oft tagebuchähnliche, ausführliche Beschreibungen des Alltags, jedoch keine klar definierten Ziele oder Berichte. Entsprechend gilt zu prüfen, ob die Dokumentation den Vorgaben der Institution entspricht.

Mängel in der Dokumentationsqualität

- Unklare Abkürzungen: Begriffe zu *KL*, *Bewohner*, *Funktion* und *Rollen* sind nicht harmonisiert.
- Falsche Zuweisungen: Der Eintrag vom 30. April 2024 gehört in ein anderes Dossier
- Vitalzeichen-Einträge wie am 19. Januar 2024 enthalten Vermutungen, die ohne Validierung stehen bleiben. Durch eine zweite Messung hätte die Vermutung bestätigt oder widerlegt werden können. Ein solcher Eintrag fehlt aber.

Dokumentationsumfang und Relevanz

- Manche Einträge sind sehr ausführlich, z. B. die Funktionsanleitung eines Hörgeräts, deren Nutzen unklar ist.
- Das Journal konzentriert sich stark auf die Beschreibung der Alltagsbewältigung; übergeordnete Ziele oder strategische Massnahmen fehlen.

Struktur und Zweck des Journals

- Es fehlt eine klare Definition, ob das Journal primär für die Altagsdokumentation oder als Grundlage für Berichte dient.
- Es gibt Diskrepanzen zwischen *Alltagsrelevanz* und *Berichtsrelevanz*. Informationen über andere Klienten können für den Alltag wichtig sein, sollten jedoch für Berichte anonymisiert oder ausgeschlossen werden.

Verbesserungspotenziale

- Anstatt Namen Rollen schreiben. Beispiel: Anstatt *Stefanie Herren* (anonymisiert) *Mutter* und den Kontakt der Mutter im Modul *Systeme* verknüpfen.
- Standardisierung der Sprache: Einheitlicher Einsatz von Hochdeutsch und Vermeidung von informellem Stil.
- Datenschutzrichtlinien: Klare Vorgaben für den Umgang mit sensiblen Daten und externen Inhalten.
- Kategorisierung und Strukturierung: Einführung klarer Kategorien und Markierungen und deren strikte Einhaltung. Allgemein gehaltene Kategorien wie *01 Beobachtung*, *Information* oder *Alltag* sind zu vermeiden.
- Institutionelle Vorgaben: Die Institution sollte definieren, welchen Zweck das Journal erfüllt und wie es genutzt werden soll.
- Technische Unterstützung: Ergänzung des Journals um Funktionen wie *Beteiligte* und Markierungen für Berichts- oder Alltagsrelevanz.

16.5 Qualitative Datenanalyse von Dossier K-2021-00079

Das Dossier K-2021-00079 zeigt spezifische Stärken und Schwächen in der Dokumentation, die nachfolgend thematisch zusammengefasst werden:

Fehlerhafte Schreibweisen und uneinheitliche Sprache

- Der Name des Klienten wird mehrfach falsch geschrieben, was die Anonymisierung erschwert.
- Vermehrte Nutzung von Dialekt und informeller Sprache, z. B. *vollgepinkelt*, *brav* oder *Schätzeli*.
- Einzelne Einträge enthalten Rechtschreibfehler, ähnlich wie in anderen Dossiers.
- Verwendung von umgangssprachlichen oder missverständlichen Begriffen wie *Hinrichtung* (27. November 2024) und *klaut* (15. Februar 2024).

Datenschutz und externe Inhalte

- Namen von Mitarbeitenden, anderen Klientinnen, externen Fachpersonen und Institutionen werden im Journal dokumentiert.
- Konkrete Adressen und Telefonnummern, z. B. eines Hofladens (16. April 2024) oder eines Lieferanten (7. November 2024), erscheinen im Journal.

Unklare und uneinheitliche Kategorisierung

- Zu viele Details in den Einträgen: z. B. detaillierte Beschreibungen von Einkäufen (5. November 2024) oder alltäglichen Erlebnissen ohne erkennbaren Zusammenhang mit dem Klienten (28. Mai 2024).
- Vermischung von pflegerischen Informationen (z. B. Medikamente, Körpertemperatur) im Journal und im Pflege-Modul. Ein Verweis im Journal würde ausreichen.
- Unterschiedliche Begriffe für ähnliche Rollen, z. B. *Kunde* und *Klient* (7. August und 19. Januar 2024).

Dokumentationsqualität

- Einige Einträge sind reflexiv und detailliert (z. B. 4. Juli 2024), was positive Ansätze erkennen lässt.
- Vermehrte Nutzung von Abkürzungen, teils missverständlich (*nie* als Abkürzung für Mitarbeitende könnte als Verneinung interpretiert werden).
- Widersprüchliche Einträge: Empfehlungen zur Senkung der Cholesterinwerte (27. März 2024) stehen im Gegensatz zu Einträgen über den Konsum von Süssgetränken und Pizza.
- Verzicht auf konkrete Ergebnisse oder Massnahmen nach bestimmten Ereignissen, z. B. fehlende Informationen nach Blutentnahme (26. November 2024).

Verbesserung der Sprache und des Stils

- Begriffe wie *brav* könnten durch neutralere Formulierungen ersetzt werden, z. B. *Der Klient folgte den Anweisungen*.
- Statt *vollgepinkelt* könnte die Formulierung *Am Boden konnte Urin festgestellt werden* angemessener sein.
- Vermeidung von Grossbuchstaben oder Ausrufezeichen zur Hervorhebung (z. B. *VOR DEM MITTAG!* am 24. März 2024).

Struktur und Zweck des Journals

- Positiv: Vermehrte Nutzung von Rollen statt Namen (*Klient*, *Betreuer*), was eine Verbesserung gegenüber anderen Dossiers darstellt.
- Diskrepanz zwischen *Alltagsrelevanz* und *Berichtsrelevanz*: Manche Einträge sind alltagsbezogen und für Berichte nicht notwendig.
- Es fehlen Richtlinien, ob das Journal primär für die Dokumentation des Alltags oder als Grundlage für Berichte dient.

Verbesserungspotenziale

- Einheitlicher Einsatz von Hochdeutsch, Vermeidung von Dialekt und informeller Sprache.
- Klare Vorgaben zum Umgang mit personenbezogenen Daten und externen Informationen (z. B. Namen und Adressen).
- Einführung klar definierter Kategorien für Berichte und Alltagsrelevanz.
- Verbesserung der Dokumentationsqualität durch Schulungen in Sprache, Datenschutz und professionellen Standards.
- Ergänzung des Journals um Funktionen wie *Beteiligte* oder Markierungen für berichts- und alltagsrelevante Einträge.
- Klärung des Zwecks und der Nutzung des Journals.

Das Dossier zeigt, dass trotz einiger positiver Ansätze (z. B. Nutzung von Rollen) Anpassungen auf sprachlicher, datenschutzrechtlicher und struktureller Ebene notwendig sind, um die Qualität der Dokumentation zu optimieren und den institutionellen Anforderungen gerecht zu werden.

17 Selbständigkeitserklärung und Nutzung von KI

Selbständigkeitserklärung

Ich bestätige, dass ich die vorliegende Arbeit selbstständig und ohne Benutzung anderer als der im Literaturverzeichnis angegebenen Quellen und Hilfsmittel angefertigt habe. Sämtliche Textstellen, die nicht von mir stammen, sind als Zitate gekennzeichnet und mit dem genauen Hinweis auf ihre Herkunft versehen.

Nutzung von Künstlicher Intelligenz

In dieser Masterthesis wurde ChatGPT-4o von OpenAI intensiv genutzt. Die Verwendung erfolgte zur Unterstützung bei der Recherche von Fachliteratur und dem Aufbau eines vertieften technischen Verständnisses. Zusätzlich lieferte das Tool Vorschläge zur sprachlichen Überarbeitung und zur Optimierung der Verständlichkeit von Textpassagen sowie zur Strukturierung und Gliederung der Arbeit.

Für diese Zwecke kamen zahlreiche Eingaben und Fragen (Prompts) zum Einsatz, wie beispielsweise zur Literaturrecherche: *«nenne Fachliteratur und Fachartikel für RAG und LLM»*, *«technische Unterstützung beim Einsatz von Python»* oder *«erstelle eine Tabelle von LLMs und den jeweiligen Nutzungsbedingungen»*.

Zur Prüfung eines stringenten Aufbaus und der Nachvollziehbarkeit der Arbeit wurden Fragen verwendet wie *«Ist der Aufbau der Arbeit stringent?»*, *«Fehlen wichtige technische Inhalte?»*, *«Entsprechen die Quellen den Standards wissenschaftlichen Arbeitens?»* oder *«Sind Kapitel harmonisch aufgebaut?»*.

Für die sprachliche und redaktionelle Überarbeitung wurden Prompts formuliert wie *«Verbessere die Verständlichkeit des folgenden Abschnitts»*, *«Formuliere prägnanter und präziser»* oder *«Prüfe die Grammatik des folgenden Abschnitts»*.

Ich bestätige, dass ich diese Studienarbeit eigenständig erstellt und KI-generierte Texte bzw. Textfragmente nicht unreflektiert übernommen habe. Zudem wurden alle Quellen verifiziert und im Literaturverzeichnis aufgeführt.

Münsingen im März 2025

Unterschrift

A handwritten signature in blue ink, consisting of a stylized, flowing script that starts with a large loop and ends with a horizontal stroke.