
Themenheft 72: Mixed Methods im Spannungsfeld. Weiterentwicklung digitaler Forschungsmethoden zwischen KI, NLP, Big Data und transdisziplinärer Forschung.
Herausgegeben von Jana Heinz, Andrea Hense und Christian Janßen

Soziale Netzwerkinderviews mit einem KI-Chatbot

Potenziale und Grenzen für die Mixed-Methods-Forschung

Marco Galle¹  und Nuria Lara Steiner² 

¹ Pädagogische Hochschule Luzern

² Universität Zürich

Zusammenfassung

Large Language Models (LLM) eröffnen neue Möglichkeiten für die qualitative Datenerhebung, insbesondere da qualitative Daten in grösserem Umfang erfasst werden können. Gleichzeitig ist zu prüfen, inwiefern LLM-Interviews vertrauensbasierte Interaktionssituationen herstellen und für kontextsensitives Fallverstehen geeignet sind. Im Beitrag wird ein LLM-basiertes soziales Netzwerkinderview vergleichend zu Peer-Gesprächen zwischen Studierenden über persönliche Netzwerkkarten untersucht. Das Design folgt einem Conversion-Mixed-Methods-Ansatz: Interview- und Gesprächsdaten werden kodiert, quantifiziert und statistisch ausgewertet. Analysiert werden (1) Nachfrageverhalten, (2) Antwortlänge, (3) Deckungsgleichheit zwischen Netzwerkkarte und Interview/Gespräch sowie (4) subjektives Erleben der Erhebungsformate. Ergebnisse zeigen, dass das LLM-Interview signifikant mehr Nachfragen erzeugt als Peer-Gespräche, ohne dass die Antwortlänge zunimmt. In beiden Formaten bleiben Akteur:innen aus der Netzwerkkarte teilweise unbesprochen; in Peer-Gesprächen werden jedoch häufiger zusätzliche Akteur:innen genannt, die nicht in der Netzwerkkarte enthalten sind. Die Studierenden schätzen das Frageverhalten im LLM-Interview als strukturiert und reflexionsfördernd, aber auch als repetitiv und teilweise kognitiv überfordernd ein. Es werden Implikationen für den Einsatz LLM-Interview und Potenziale für Conversion-Mixed-Methods-Analysen diskutiert.

Social Network Interviews with an AI Chatbot. Potentials and Limitations for Mixed Methods Research

Abstract

Large Language Models (LLMs) open up new possibilities for qualitative data collection, particularly by enabling the collection of qualitative data at greater scale. At the same time, it is necessary to critically examine whether LLM-based interviews can establish

trust-based interactional situations and whether they are suitable for context-sensitive, case-oriented understanding. This paper comparatively investigates an LLM-based social network interview and peer conversations between students about personal network maps. The study follows a conversion mixed-methods design: interview and conversation data are coded, quantified, and analysed statistically. The analysis focuses on (1) follow-up questioning behaviour, (2) response length, (3) congruence between network maps and interviews/conversations, and (4) participants' subjective experiences of the data collection formats. The results show that the LLM-based interview generates significantly more follow-up questions than peer conversations, without an increase in response length. In both formats, some actors depicted in the network maps remain unaddressed; however, peer conversations more frequently elicit additional actors who are not included in the network maps. Student teachers perceive the questioning behaviour in the LLM-based interview as structured and conducive to reflection, but also as repetitive and, at times, cognitively demanding. The paper discusses implications for the use of LLM-based semi-structured interviews and outlines potentials for conversion mixed-methods analyses.

1. Einleitung

Large Language Models (LLMs), im wissenschaftlichen und öffentlichen Diskurs häufig als (generative) künstliche Intelligenz bezeichnet, sind neuronale Sprachmodelle, die auf umfangreichen Textkorpora trainiert werden und natürliche Sprache probabilistisch erzeugen und verarbeiten. Aufgrund ihrer Fähigkeit, syntaktische, semantische und pragmatische Muster menschlicher Kommunikation zu imitieren (Han u. a. 2024), eröffnen sie neue Perspektiven für qualitative Datenerhebung und Mixed-Methods-Forschung. Es wird erwartet, dass dialogfähige LLMs qualitative Daten in grösserem Umfang und bei grösseren Stichproben generieren können als klassische Interviewdesigns. Damit eröffnen sich Potenziale für Conversion-Mixed-Methods-Designs, in denen qualitative Daten quantifiziert und aus mehreren methodischen Perspektiven analysiert werden (Tashakkori u. a. 2021).

Allerdings ist aus einer methodologischen Perspektive zu prüfen, ob erstens LLM-Interviews vertrauensbasierte Interaktionssituationen herstellen können, die eine genuine Selbstöffnung der Befragten ermöglichen (Wuttke u. a. 2025). Zweitens ist zu untersuchen, inwiefern LLM-Interviews den Kriterien von Menschen durchgeführte Leitfadenterviews, namentlich Offenheit, Spezifität, Kontextualität und Relevanz (Helfferich 2014; Przyborski und Wohlrab-Sahr 2021), entsprechen.

Im Beitrag werden diese Fragen im Kontext sozialer Netzwerkinterviews untersucht (Bellotti 2015). Ziel ist es, ein LLM-Chatbot als interviewenden Akteur zu erproben: Studierende erstellten zunächst eine persönliche Netzwerkkarte und wurden anschliessend entweder von einem LLM-Chatbot interviewt oder besprachen ihre Karte mit einer Mitstudentin bzw. einem Mitstudenten. Die Chatbotprotokolle und die transkribierten

Gespräche wurden entlang von vier Themenbereichen verglichen: (1) Nachfrageverhalten, (2) Länge der Antworten, (3) Deckungsgleichheit zwischen Netzwerkkarten und LLM-Interviews respektive Peer-Gesprächen sowie (4) subjektives Erleben der Studierenden.

2. Methodologische Grundlagen

2.1 *Mixed Methods in der Datenerhebung der sozialen Netzwerkforschung*

In Mixed-Methods-Designs werden qualitative und quantitative Verfahren mit dem Ziel kombiniert, einen Untersuchungsgegenstand aus unterschiedlichen methodischen Perspektiven zu erfassen und auszuwerten (Creswell und Plano Clark 2018). Da im vorliegenden Beitrag die Datenerhebung im Zentrum steht, liegt der Fokus auf der Integration beider Datentypen bereits im Erhebungsprozess beziehungsweise in einem Erhebungsinstrument. Dies ermöglicht es, einerseits fallbasierte, kontextsensitive Daten zu erheben, andererseits eine grosse Anzahl von Personen zu befragen.

In der sozialen Netzwerkforschung sind Mixed-Methods-Designs bislang selten vertreten (Fröhlich, Rehm, und Rienties 2020). Dominant sind monomethodische Zugänge: Qualitative Netzwerkinderviews liefern narrativ eingebettete Deutungen relationaler Strukturen, sind jedoch aufgrund des Erhebungs- und Auswertungsaufwands auf kleine Fallzahlen beschränkt. Quantitative Netzwerkstudien hingegen erfassen relationale Strukturen grösserer Gruppen, verzichten jedoch infolge der Standardisierung weitgehend auf subjektive Erfahrungs- und Erlebensbeschreibungen sozialer Beziehungen. In Mixed-Methods-Studien wird versucht, diese Begrenzungen zu überwinden, indem qualitative und quantitative Daten entweder sequenziell oder parallel erhoben werden (Fröhlich 2023).

Weniger diskutiert sind bislang Erhebungsinstrumente, die qualitative und quantitative Daten innerhalb eines einzigen Instruments erfassen. Solche integrierten Erhebungsformate stärken die Verbindung beider Datentypen, da sie zeitlich unmittelbar miteinander verknüpft erhoben werden. Dies ist insbesondere im Kontext sozialer Netzwerke relevant, da sich relationale Strukturen dynamisch verändern und bei zeitlichen Verzögerungen zwischen Erhebungsphasen nicht gewährleistet werden kann, dass dieselben relationalen Strukturen erfasst werden (McCormick u. a. 2011; Stegbauer 2025).

Ein Beispiel für ein solch integriertes Erhebungsformat stellt das im SNF-Projekt «DiaMaNt» (Kreis und Brunner 2022) entwickelte Instrument zur Erfassung sozialer Netzwerke von Lehramtsstudierenden in bedeutsamen Lernsituationen des Praktikums dar. Die Studierenden zeichneten zunächst in einer standardisierten PowerPoint-Datei eine persönliche Netzwerkkarte, in der sie für ihre Kompetenzentwicklung bedeutsame Akteur:innen aufführten. Anschliessend besprachen sie diese Netzwerkkarte mit einer Mitstudentin oder einem Mitstudenten. Die Gesprächspartner:innen erhielten den

Auftrag, verständnissichernde und vertiefende Nachfragen zu stellen, und die Gespräche wurden aufgezeichnet. Ziel war es, zusätzliche Kontextinformationen zu generieren, um die Netzwerkkarten fundierter interpretieren und reliabler kodieren zu können (Galle, Kreis, und Hiebler 2022). Trotz eines insgesamt informationsreichen Datensatzes zeigte sich eine Einschränkung dieses Erhebungsformats: Vertiefende Nachfragen wurden nur selten gestellt. Lediglich 12.27 % der Studierenden (n = 24) formulierten Rückfragen, von denen wiederum nur ein Teil inhaltlich relevant war. Zudem zeigten sich Inkonsistenzen zwischen der Netzwerkkarte und dem Gespräch: Bei 32.28 % der Studierenden wurden Akteur:innen entweder nur in der Netzwerkkarte oder nur im Gespräch genannt (Galle, Kreis, und Hiebler 2025).

Zwar boten die Peer-Gespräche einen vertieften Einblick in die Netzwerkkarten der Studierenden, jedoch ist das Potenzial der Datenerhebung mit den wenigen Nachfragen noch nicht ausgeschöpft. An dieser Stelle setzt dieser Beitrag an, halbstrukturierte, leitfadengestützte Einzelinterviews mit LLMs durchzuführen, um mit qualitätsvollen Nachfragen vertiefte Informationen zu den Netzwerkkarten zu erhalten.

2.2 Halbstrukturierte und leitfadengestützte Einzelinterviews

Halbstrukturierte Leitfadeninterviews sind zwischen offenen und stark strukturierten Interviewformen verortet (Loosen 2016). Kennzeichnend ist ein thematisch strukturierter Leitfaden, der aus dem Forschungsinteresse abgeleitet wird und als Orientierungsrahmen dient. Der Leitfaden legt fest, was thematisiert werden soll, nicht jedoch die genaue Reihenfolge oder Formulierung der Fragen (Przyborski und Wohlrab-Sahr 2021; Kruse 2015). Ziel halbstrukturierter Leitfadeninterviews ist es, Strukturierungs- und Offenheitsanforderungen auszubalancieren. Der Leitfaden stellt sicher, dass zentrale Aspekte des Forschungsinteresses in allen Interviews vergleichbar adressiert werden, während zugleich Raum für subjektiv bedeutsame Themen und individuelle Akzentuierungen bleibt (Helfferich 2014).

Für die Qualität leitfadengestützter Interviews lassen sich vier zentrale Gestaltungskriterien benennen: Offenheit, Spezifität, Kontextualität und Relevanz (Przyborski und Wohlrab-Sahr 2021; Helfferich 2011). *Offenheit* bezieht sich insbesondere auf erzählgenerierende Einstiegsfragen, die es den interviewten Personen ermöglichen, den Gegenstand aus ihrer Perspektive darzustellen. *Spezifität* meint vertiefende Nachfragen, die an bereits Angedeutetes anknüpfen und zur Präzisierung beitragen. *Kontextualität* zielt darauf ab, Aussagen in ihrer situativen, sozialen, institutionellen und biographischen Einbettung sichtbar zu machen und deren subjektive Bedeutung zu erschliessen. *Relevanz* meint, dass das Gespräch über für den Untersuchungsgegenstand und für die interviewte Person relevante Inhalte geführt wird. Falls nicht relevante Inhalte erzählt werden, wird das Interview behutsam zum Gegenstand zurückgeführt (Kruse 2015; Helfferich 2014).

Für diese Interviewform sind kontext- und fallbezogene Nachfragen bedeutsam. Ein Nachfragetyp ist für das in diesem Beitrag vorgestellte Erhebungsinstrument relevant: Immanente Nachfragen sind Fragen, die unmittelbar an vorherige Äußerungen der interviewten Person anschliessen. Sie dienen dazu, implizite, unklare oder nur angedeutete Aspekte zu explizieren (Przyborski und Wohlrab-Sahr 2021).

2.3 Funktion von Large Language Models (LLMs) in Interviews

Im Diskurs zur Nutzung von LLMs in der qualitativen Datenerhebung lassen sich zwei Positionen herausarbeiten: Erstens fungiert das LLM als unterstützende Instanz im Hintergrund eines von Menschen geführten Interviews, etwa durch Vorschläge für Nachfragen, Fokussierungen oder Zusammenfassungen (Liu u. a. 2025; Zhang u. a. 2025). Zweitens übernimmt das LLM selbst die Rolle der Interviewinstanz und führt das Gespräch direkt mit den Befragten (Wuttke u. a. 2025).

Für eine *unterstützende Rolle von LLMs (erste Position)* und damit gegen die zweite Position spricht, dass LLMs in der Lage sind, sprachlich kohärente Inhalte zu generieren und damit im Interview vage oder unvollständige Antworten zu präzisieren und übersehene Aspekte nachträglich zu adressieren (Zhang u. a. 2025). Allerdings verfügen LLMs weder über genuines kritisches Denken noch über ein eigenständiges Verständnis sozialer Kontexte. Sie können keine genuine zwischenmenschliche Interaktion führen und keine auf die spezifischen Situationen der interviewten Personen bezogene Reflexion oder Kontextdeutung leisten; vielmehr folgen sie den bereitgestellten Eingaben und generieren darauf basierende Antworten (Burleigh und Wilson 2024; Williams und Ingleby 2025). Darüber hinaus besteht das Risiko, dass fehlerhafte, stereotype oder voreingenommene Antworten produziert werden, die sich u. a. auf Trainingsdaten und algorithmische Gewichtung zurückführen lassen (Han u. a. 2024).

Wenn *LLMs selbst das Interview führen (zweite Position)*, können geografisch weitverbreitete Personen interviewt werden, was zu einer Erhöhung der Stichprobendiversität beitragen kann. Zudem wird die Flexibilität der Interviewplanung erhöht, da die Durchführung des Interviews von den befragten Personen unabhängig vom Forschungsteam selbst bestimmt werden kann (Williams und Ingleby 2025). Ebenso lassen sich mit LLM-Interviews sehr viele Personen innert kürzester Zeit interviewen, was eine Mixed-Methods-Analyse unterstützt und die Generalisierbarkeit der Ergebnisse erhöht (Wuttke u. a. 2025). Zur Entwicklung eines interviewführenden LLMs schlagen Chopra und Haaland (2023) einen modularen Aufbau bestehend aus vier Funktionen vor: Prüfung der Passung von Antworten zu den Forschungsfragen, kontextuelle Einbettung neuer Fragen inklusive Zusammenfassungen, Identifikation bislang unadressierter leitfaden- oder theoriebezogener Inhalte sowie Möglichkeit, Fragen flexibel zu überspringen. Zudem unterstreichen sie die Notwendigkeit einer engen Abstimmung mit Ethikkommissionen bei der Entwicklung und Implementierung solcher Technologien.

3. Forschungsfragen

Zum Einsatz von Large Language Models (LLMs) in der qualitativen Datenerhebung liegen erste Erfahrungen und Studien vor. Weitere Studien zu unterschiedlichen Einsatzmöglichkeiten sind notwendig, um deren Potenziale und Grenzen evidenzbasiert besser einschätzen zu können. LLMs bieten in der Mixed-Methods-Forschung neue Möglichkeiten, qualitative und quantitative Datenerhebungsformate zu verknüpfen. Im Kontext sozialer Netzwerkforschung ist die zeitgleiche Erhebung von qualitativen und quantitativen Daten vielversprechend, da relationale Strukturen sich im zeitlichen Verlauf ändern können und eine zeitlich verschobene Erhebung von quantitativen und qualitativen Daten möglicherweise unterschiedliche Netzwerkstrukturen erfasst.

Für diesen Beitrag wurde ein Erhebungsinstrument entwickelt, in dem zeitgleich qualitative und quantitative Daten zu persönlichen Netzwerken von Lehramtsstudierenden erhoben wurden. Es wird untersucht, welche Gemeinsamkeiten und Unterschiede sich zwischen einem schriftlich geführten, LLM-basierten sozialen Netzwerkinterview (Chatbot) und einem Gespräch zwischen Studierenden (Peer-Gespräch) zu ihren persönlichen Netzwerkkarten ergeben.

Es werden folgende Forschungsfragen beantwortet:

1. *Nachfrageverhalten*: Wie viele Nachfragen stellt der Chatbot und unterscheidet sich die Anzahl der Nachfragen zwischen LLM-Interviews und Peer-Gesprächen? Es wird erwartet, dass der Chatbot aufgrund seiner Fähigkeit zur sprachlichen Imitation und systematischen Gesprächsführung mehr Nachfragen stellt als Studierende in Peer-Gesprächen.
2. *Antwortlänge*: Wie umfangreich sind die Antworten der Studierenden im LLM-Interview und unterscheiden sich die Antwortlängen (Zeichenanzahl) zwischen LLM-Interviews und Gesprächen zwischen Studierenden? Aufgrund der erwarteten qualitativ besseren und stärker fokussierten Nachfragen des Chatbots wird angenommen, dass die Antworten in LLM-Interviews länger ausfallen.
3. *Deckungsgleichheit von Netzwerkkarte und LLM-Interview/Peer-Gespräch*: Wie deckungsgleich sind Netzwerkkarten und LLM-Interviews bzw. Gespräche hinsichtlich der erwähnten Akteursgruppen (z. B. Akteur:innen nur in der Karte, nur im Gespräch oder in beiden Datenquellen)? Da der Chatbot in der Lage ist, die Regel, nach allen in der Netzwerkkarte genannten Personen nachzufragen, stringent einzuhalten, wird eine höhere Deckungsgleichheit im LLM-Interview erwartet als in Peer-Gesprächen.
4. *Subjektives Erleben der Erhebungsformate*: Welche positiven und negativen Aspekte benennen Studierende in Bezug auf die Nutzung des Chatbots respektive des Peer-Gesprächs?

4. Methode

4.1 *Beschreibung des SNF-Projekts «network.»: Ausgangslage und Forschungsinteressen*

Kooperation gilt als zentrale professionelle Handlungskompetenz von Lehrpersonen (Fabel-Lamla und Gräsel 2020), z. B. beim Teamteaching (York-Barr, Ghere, und Sommerness 2007) oder im Kontext inklusiver Pädagogik (Grosche und Moser Opitz 2023). Empirische Befunde zeigen, dass kooperative Praktiken positiv auf das Lernen der Schüler:innen wirken (Ronfeldt 2015; Saka 2021). Lehramtsstudierende erleben jedoch vielfach eine Diskrepanz zwischen diesen Anforderungen und ihrer eigenen Vorbereitung: Trotz curricularer Verankerung (Galle 2026) berichten sie, nur unzureichend auf kooperative Tätigkeiten vorbereitet zu sein (Keller-Schneider und Hericks 2020; Winkler und Freisler-Mühlemann 2023). In der Lehrpersonenbildung ist das Thema «Kooperation lernen» bisher wenig beforscht (Bauer und Fabel-Lamla 2020).

Im vom SNF geförderten Forschungsprojekt «network.» werden Lerngelegenheiten von Studierenden aus einer soziokonstruktivistischen Perspektive (Resnick, Levine und Teasley 1991) untersucht. Lernen wird als ein sozial eingebetteter und interaktiver Prozess definiert. Grundlage bildet die Activity Theory (Engeström 1999), die in der Kompetenzentwicklung als Ergebnis der Wechselwirkung zwischen Akteur:innen (Subjekte), Instrumenten (Artefakte) und dem Gegenstand (Objekt) beschrieben wird (Kreis und Brunner 2022). Die Vielzahl von Akteur:innen, mit denen Studierende im Studium interagieren und von denen sie etwas bezüglich lernzentrierter Kooperation lernen können, lässt sich als soziales Netzwerk fassen (McCormick u. a. 2011). Eine soziale Netzwerkperspektive fokussiert die relationalen Strukturen der Studierenden und deren Wechselwirkung mit ihren Lerntätigkeiten (Gruber, Hirschmann, und Rehl 2018).

Neben inhaltlichen Fragestellungen zu den Lerngelegenheiten von Studierenden interessiert auch die methodische Exploration der Möglichkeiten von LLM-Interviews. Auf dieses Thema wird nachfolgend eingegangen.

4.2 *Forschungsdesign*

Das Forschungsdesign folgt einem Conversion-Mixed-Methods-Design, bei dem qualitative Daten in quantitative Variablen transformiert und anschliessend statistisch ausgewertet werden (Tashakkori, Johnson, und Teddlie 2021). Im ersten Schritt werden qualitative, nicht standardisierte Interview- und Gesprächsdaten oder Antworten auf offene Fragen erhoben und kodiert. In einem zweiten Schritt erfolgt deren Quantifizierung, um über die ausgezählten Kodierungen zu berichten (Forschungsfrage 1 bis 4) und statistische Analysen durchzuführen (Forschungsfrage 1 bis 3).

4.3 Stichprobe

Die Erhebung wurde 2025 mit Lehramtsstudierenden der Sekundarstufe 1 an der Pädagogischen Hochschule Luzern durchgeführt, die sich im vorletzten Mastersemester befanden. Insgesamt wurden persönliche Netzwerkkarten von 71 Studierenden erhoben (Rücklauf = 69%). Zusätzlich liegen 56 LLM-Interviews vor (Rücklauf = 88%) sowie 15 Audioaufzeichnungen aus Peer-Gesprächen (Rücklauf = 37%). Der vergleichsweise geringe Rücklauf bei den Audiodaten ist auf technische Störungen der eingesetzten Webapplikation zurückzuführen, insbesondere auf nicht startende oder nicht erfolgreich hochgeladene Audioaufnahmen.

Die Stichprobe besteht zu 53.3% aus weiblichen und zu 46.7% aus männlichen Studierenden. Das Durchschnittsalter beträgt 26.32 Jahre (SD = 5.23).

4.4 Instrument der Datenerhebung

Die 45-minütige Datenerhebung erfolgte in einer Lehrveranstaltung. Die Erhebung wurde mithilfe einer Webapplikation durchgeführt und folgte einem mehrstufigen Ablauf (Abbildung 1).

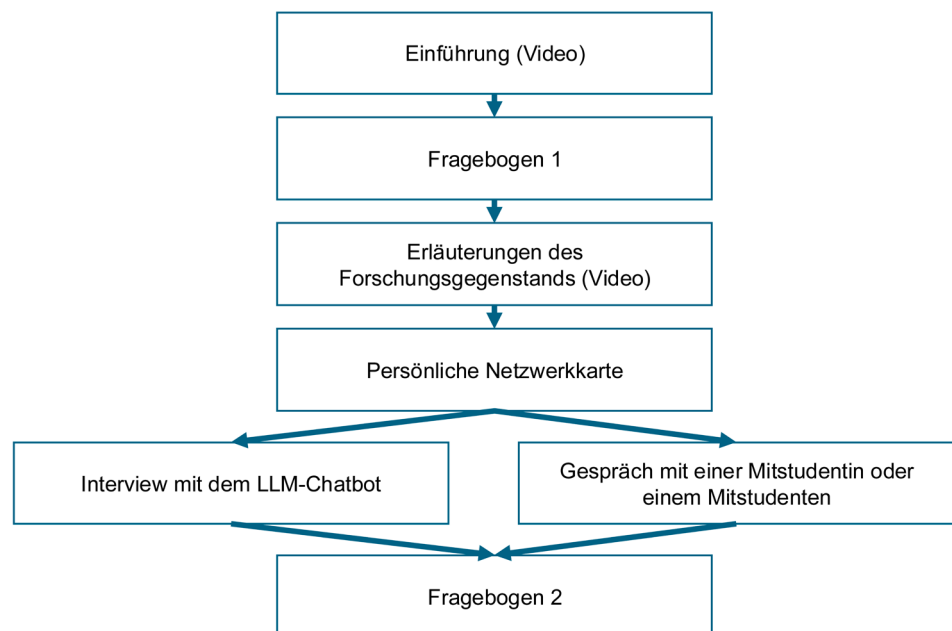


Abb. 1: Schritte der Webapplikation (Galle, Strässle, und Loser 2025).

1. Nach dem Öffnen der Webapplikation erhielten die Studierenden über ein Video allgemeine Informationen zum Forschungsprojekt sowie zur Verwendung der erhobenen Daten.
2. Anschliessend beantworteten die Studierenden einen Fragebogen zu personenbezogenen Merkmalen sowie zu Lerngelegenheiten in Bezug auf den Untersuchungsgegenstand.
3. Im dritten Schritt sahen sich die Studierenden ein Erklärvideo zum Untersuchungsgegenstand an, in dem zentrale Aspekte lernzentrierter Kooperation und sozialer Netzwerke thematisiert wurden.
4. Darauf aufbauend erstellten die Studierenden eine persönliche Netzwerkkarte (Hollstein, Töpfer, und Pfeffer 2020; Schönhuth und Gamper 2013). Hierzu benannten sie Akteur:innen, von denen sie in den vergangenen zwölf Monaten etwas zum Thema lernzentrierte Kooperation gelernt hatten. Für jede Akteurin bzw. jeden Akteur gaben sie zusätzlich an, in welchem Setting die Lernsituation stattfand, was sie für ihre zukünftige berufliche Tätigkeit gelernt hatten und wie bedeutsam die jeweilige Lerngelegenheit für ihre Kompetenzentwicklung eingeschätzt wurde.
5. Im nächsten Schritt erhielten die Studierenden entweder den Auftrag, sich von dem Chatbot zu ihrer Netzwerkkarte interviewen zu lassen oder sich über ihre Netzwerkkarte mit einer Mitstudentin bzw. einem Mitstudenten auszutauschen. Für das Gespräch erhielten die Studierenden explizit den Auftrag, Rückfragen zu stellen. Die Zuordnung zu den Erhebungsformaten erfolgte zufällig auf Ebene der Mentoratsgruppen, sodass alle Studierenden einer Gruppe denselben Auftrag erhielten.
6. Abschliessend beantworteten die Studierenden Evaluationsfragen zum Erleben des jeweiligen Erhebungsformats. Für den vorliegenden Beitrag werden zwei offene Fragen berücksichtigt: *Was hat Ihnen bei der Interaktion [mit dem Chatbot / im Gespräch mit Ihrer Mitstudentin bzw. Ihrem Mitstudenten] gefallen? Was hat Ihnen bei der Interaktion [mit dem Chatbot / im Gespräch mit Ihrer Mitstudentin bzw. Ihrem Mitstudenten] nicht gefallen?*

Die Studierenden wurden durch videobasierte und textuelle Erläuterungen systematisch durch alle Schritte geführt. Inhaltliche Rückfragen an die anwesenden Mentorierenden traten nicht auf; Unterstützung erfolgte ausschliesslich bei technischen Problemen (z. B. Ladefehler oder fehlgeschlagene Audioaufnahmen).

Die Interviewdaten des Chatbots wurden temporär über die API von OpenAI verarbeitet (Serverstandort: Schweiz) und nach der Speicherung auf universitären Servern gelöscht. Die anonymisierten Forschungsdaten werden auf SWISSUbase veröffentlicht, sobald alle Hauptpublikationen akzeptiert sind. Der Quellcode des Chatbots sowie eine Dokumentation der verwendeten Prompts sind auf GitHub open source verfügbar (Strässle 2025).

4.5 Entwicklung des Chatbots

Die Entwicklung des Chatbots orientierte sich an den von Chopra und Haaland (2023) formulierten Grundprinzipien für LLM-Interviews. Diese umfassen (1) die Prüfung der Passung von Antworten zu den Forschungszielen, (2) die Generierung kohärenter und kontextuell eingebetteter Folgefragen einschliesslich Paraphrasierungen und Zusammenfassungen, (3) die Identifikation bislang nicht thematisierter, leitfadenrelevanter Inhalte sowie (4) die Möglichkeit für Teilnehmende, Fragen zu überspringen.

Die technische Umsetzung erfolgte in Zusammenarbeit mit der Firma smartive AG in einem mehrstufigen Entwicklungsprozess. Ausgangspunkt bildete ein vortrainiertes Sprachmodell, das durch überwachtes Lernen und Reinforcement Learning aus menschlichem Feedback hinsichtlich seiner dialogischen Fähigkeiten optimiert wurde (von OpenAI verantwortet). Darauf aufbauend wurde ein Instruction Fine-Tuning durchgeführt, um das Modell gezielt auf die Anforderungen des Projekts und interviewbezogene Anwendungsszenarien zuzuschneiden. In einem weiteren Schritt wurden die eingesetzten Prompts iterativ getestet und angepasst, um Relevanz, Verständlichkeit und Passung der Nachfragen zu verbessern. Ergänzend wurde eine Retrieval-Augmented-Generation-Funktion integriert, die den Zugriff auf eine projektspezifische Wissensbasis ermöglicht, ohne erneutes Modelltraining zu erfordern. Zur Sicherstellung stabiler und fokussierter Interaktionen wurde ein modular aufgebautes System mit drei spezialisierten LLM-Komponenten implementiert (Abbildung 2).

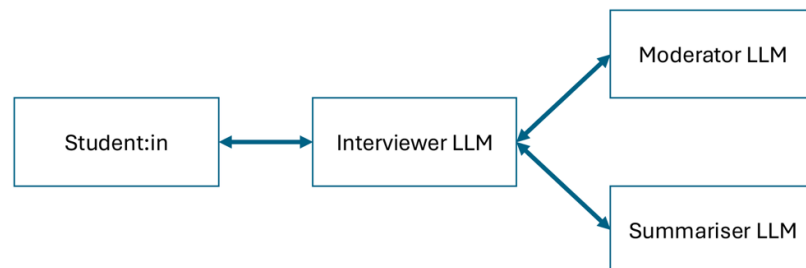


Abb. 2: Modularer Aufbau des Chatbots (Galle, Strässle, und Loser 2025)

Das Interviewer-LLM (GPT-4o) führte das Gespräch, bezog sich auf die jeweilige Netzwerkkarte und generierte kontextbezogene Folgefragen. Unterstützt wurde es durch ein Moderation-LLM (GPT-4o mini), das unter anderem Angemessenheit der Fragen, Einhaltung ethischer Grundprinzipien und thematische Fokussierung überwachte, sowie ein Summariser-LLM (GPT-4o mini), das relevante Gesprächsinhalte fortlaufend verdichtete, um den vollständigen Gesprächskontext verfügbar zu halten (Galle, Strässle, und Loser 2025).

Neben der technischen Entwicklung wurden methodische Handlungsanweisungen ausgearbeitet und in die Prompts integriert. Ziel des LLM-Interviews war es, vertiefende Informationen zu den Netzwerkkarten zu generieren. Dabei bedeutsam waren

immanente Nachfragen, die unmittelbar an vorherige Äusserungen der Interviewten bzw. an Elemente der Netzwerkkarte anschlossen (Przyborski und Wohlrab-Sahr 2021). Die methodische Grundlage für die Promptgestaltung bildeten die Arbeiten von Helfferich (2011; 2014). Die zentralen Interviewregeln umfassten unter anderem eine klare Rollen- und Kontextdefinition, eine standardisierte Einführung in das Interview, die Begrenzung auf eine nicht suggestive W-Frage pro Prompt, die behutsame Rückführung bei thematischen Abschweifungen sowie einen standardisierten Abschluss, bei dem nach weiteren in der Netzwerkkarte nicht sichtbaren Akteur:innen gefragt wurde (Galle, Strässle und Loser 2025).

4.6 Datenauswertung

Die Audioaufnahmen der Peer-Gespräche wurden mit noScribe (Dröge 2023) transkribiert und anschliessend manuell korrigiert. Da der inhaltliche Gehalt der Aussagen im Zentrum stand, wurde ein reduziertes Transkriptionsverfahren angewandt: Die Transkripte wurden sprachlich geglättet; auffällige Pausen wurden abhängig von ihrer Dauer mit (.) bis (...) markiert.

Im Anschluss wurden die Netzwerkkarten, Chatverläufe und Transkripte in MAXQDA 24 (VERBI Software 2024) importiert und in Bezug auf die Forschungsfragen ausgewertet. Die statistischen Analysen erfolgten in SPSS (Version 31).

- *Forschungsfrage 1:* Alle Nachfragen in den LLM-Interviews sowie in den Peer-Gesprächen wurden kodiert. Die jeweils erste Frage zu einer Person im LLM-Interview wurde nicht berücksichtigt, da sie der Gesprächseröffnung diente. Anschliessend wurde die Anzahl der Nachfragen pro Student:in ermittelt und mittels Mann-Whitney-U-Test zwischen den Gruppen Chatbot und Gespräch verglichen.
- *Forschungsfrage 2:* Die Antworten in den LLM-Interviews und in den Gesprächen wurden kodiert, wobei Zeitmarken und Sprecher:innenkürzel ausgeschlossen wurden. Die Länge der kodierten Segmente wurde in MAXQDA in Zeichen ausgegeben, nach SPSS übertragen und ebenfalls mittels Mann-Whitney-U-Test zwischen den beiden Gruppen verglichen.
- *Forschungsfrage 3:* Die in den Netzwerkkarten sowie in den LLM-Interviews bzw. Gesprächen erwähnten Akteur:innen wurden kodiert und miteinander verglichen. Erfasst wurde, wie viele Akteur:innen ausschliesslich in der Netzwerkkarte oder ausschliesslich im LLM-Interview bzw. Peer-Gespräch genannt wurden. Unterschiede zwischen den Gruppen wurden anschliessend mittels Mann-Whitney-U-Test überprüft.
- *Forschungsfrage 4:* Die Antworten auf die beiden offenen Fragebogenfragen wurden mittels strukturierender qualitativer Inhaltsanalyse ausgewertet (Kuckartz und Rädiker 2024). Das Kategoriensystem wurde induktiv entwickelt und durch eine zufriedenstellende Interkoder-Übereinstimmungsprüfung abgesichert (76.19 % Übereinstimmung; Kappa nach Brennan und Prediger [1981] = .75).

5. Ergebnisse

5.1 Anzahl Nachfragen (1. Forschungsfrage)

Die Anzahl der gestellten Nachfragen unterscheidet sich deutlich zwischen den beiden Gruppen. In Gruppe 1 (Peer-Gespräch) wurden im Durchschnitt 0.67 Nachfragen gestellt ($SD = .98$; Median = 0; $n = 15$). In Gruppe 2 (LLM-Interview) lag die durchschnittliche Anzahl der Nachfragen bei 10.43 ($SD = 5.33$; Median = 10; $n = 56$). Der Unterschied ist statistisch hochsignifikant mit starkem Effekt ($U = 10.00$; $Z = -5.79$; $p < .001$; $r = 0.68$).

5.2 Länge der Antworten (2. Forschungsfrage)

Die Analyse der Antwortlängen zeigt zwischen den beiden Gruppen keinen statistisch signifikanten Unterschied ($U = 350.00$; $Z = -0.99$; $p = .324$). In Gruppe 1 (Peer-Gespräch, $n = 15$) lag die durchschnittliche Länge der Antworten bei 1'285.07 Zeichen ($SD = 780.37$; Median = 1'242). In Gruppe 2 (LLM-Interview, $n = 56$) betrug die durchschnittliche Antwortlänge 1'407.71 Zeichen ($SD = 928.93$; Median = 1'441).

5.3 Deckungsgleichheit (3. Forschungsfrage)

Die Analyse der Deckungsgleichheit zwischen den Netzwerkkarten und den beiden Erhebungsformaten zeigt eine vollständige Übereinstimmung – das heisst, dass in der Netzwerkkarte sowie im Gespräch bzw. Interview dieselben Akteur:innen genannt wurden – in 40.00 % der Fälle der Gruppe «Peer-Gespräch» und in 48.21 % der Fälle der Gruppe «LLM-Interview».

Bezogen auf Akteur:innen, die in der Netzwerkkarte aufgeführt, jedoch im Peer-Gespräch bzw. LLM-Interview nicht thematisiert wurden, fehlten in Gruppe «Peer-Gespräch» im Mittel 0.80 Personen ($SD = 1.15$; Median = 0; $n = 15$). In der Gruppe «LLM-Interview» betrug dieser Wert im Mittel 1.29 Personen ($SD = 1.86$; Median = 0; $n = 56$). Zwischen den beiden Gruppen zeigte sich kein signifikanter Unterschied (Mann-Whitney-U-Test: $U = 381.50$; $Z = -0.59$; $p = .556$).

Hinsichtlich fehlender Akteur:innen in der Netzwerkkarte, also Akteur:innen, die im LLM-Interview oder Peer-Gespräch auftauchen, aber nicht in der ursprünglichen Karte, zeigt sich, dass bei der Gruppe «Peer-Gespräch» im Mittel 0.33 Akteur:innen fehlen ($SD = 0.49$; Median = 0; $n = 15$), bei der Gruppe «LLM-Interview» dagegen 0.04 ($SD = 0.19$; Median = 0; $n = 56$). Der Unterschied ist statistisch hochsignifikant mit mittlerem Effekt ($U = 290.00$; $Z = -3.37$; $p < .001$; $r = 0.40$). Somit wurden in Peer-Gesprächen mehr Akteur:innen genannt als im LLM-Interview, die in den Netzwerkkarten nicht erwähnt wurden.

5.4 Subjektives Erleben der Erhebungsformate (4. Fragebogen)

Da die Beschreibungen zum Erleben der beiden Erhebungsformate stark variierten, wurden für das LLM-Interview und die Peer-Gespräche jeweils eigenständige Kategoriensysteme entwickelt.

Die positiven Rückmeldungen zum Erleben des LLM-Interviews konnten in vier Kategorien geordnet werden:

1. *Gesprächsführung* (n = 25): Die Studierenden beschrieben das Interview als strukturiert, responsiv und dialogisch. Geschätzt wurde, dass Nachfragen an vorherige Aussagen anschlossen und sich konkret auf die Netzwerkkarte bezogen. Sie betonten das Gefühl, verstanden zu werden, sowie hoben die präzisen Zusammenfassungen und Rückbezüge durch den Chatbot hervor.
2. *Unterstützende Fragen zur Reflexion* (n = 13): Die Nachfragen regten die Studierenden dazu an, ihr eigenes Handeln zu reflektieren und Perspektiven zu wechseln. Der Chatbot wurde als strukturierende Unterstützung des Reflexionsprozesses wahrgenommen.
3. *Positives zur Mensch-Maschine-Interaktion* (n = 9): Die Studierenden unterstrichen die Niedrigschwelligkeit, klare Funktionsweise und gute Bedienbarkeit des Chatbots. Das Format wurde als zeitgemäss und im Vergleich zu klassischen Fragebogenverfahren als interaktiver erlebt.
4. *Sprachlich elaborierter Bot* (n = 4): Es wurde ein professioneller, freundlicher und respektvoller Tonfall attestiert.

Kritische Rückmeldungen zum LLM-Interview liessen sich in fünf Kategorien fassen:

1. *Wiederholungen ähnlich formulierter Fragen* (n = 28): Die Studierenden nahmen wiederholte oder nur minimal variierte Fragetypen als monoton wahr und beschrieben den Eindruck eines schematischen, automatisierten Frageverhaltens.
2. *Zu viele und zu detaillierte Nachfragen* (n = 19): Die Studierenden beschrieben, dass der Chatbot wiederholt bei einzelnen Punkten verharret, obwohl sie keine zusätzlichen Inhalte beisteuern konnten.
3. *Fragen zu schwierig zu beantworten* (n = 8): Zudem waren die Fragen kognitiv herausfordernd, da vom Chatbot Informationen oder Beispiele erfragt wurden, die nicht (mehr) Erinnerbar waren.
4. *Kritik an der Mensch-Maschine-Interaktion* (n = 4): Die Interaktion mit dem Chatbot wurde als künstlich, unpersönlich und datenschutzbezogen bedenklich wahrgenommen.
5. *Interaktion hat nicht zur Reflexion beigetragen* (n = 3): Es wurden keine Reflexionsprozesse angestoßen.

Bei den positiven Rückmeldungen zum Gespräch mit der Mitstudentin oder dem Mitstudenten konnten die Aussagen vier Kategorien zugeordnet werden:

1. *Gesprächssituation* (n = 11): Das Gespräch wurde als offen, ehrlich und wertschätzend beschrieben. Aktives Zuhören, respektvolle Kommunikation und eine entspannte Atmosphäre trugen zu einem positiven Erleben bei.
2. *Erkenntnisgewinn* (n = 7): Die Gespräche ermöglichten Perspektivwechsel, machten Gemeinsamkeiten und Unterschiede in den Netzwerken sichtbar und regten zur Reflexion eigener sozialer Beziehungen an.
3. *Positive Aspekte der Datenerhebung* (n = 3): Das Gespräch wurde als ergänzendes Element geschätzt, insbesondere die Möglichkeit, mündlich differenzierter zu erläutern, was schriftlich nur verkürzt darstellbar war.
4. *Austausch über Erfahrungen und Perspektiven* (n = 2): Der informelle Charakter des Erfahrungsaustauschs wurde als positiv erlebt, ohne expliziten Erkenntnisgewinn.

Kritische Rückmeldungen zum Peer-Gespräch liessen sich vier Kategorien zuordnen:

1. *Verpflichtendes Gespräch im Forschungssetting* (n = 7): Da das Gespräch als verpflichtend wahrgenommen wurde, erlebten es die Studierenden als künstlich oder erzwungen.
2. *Fehlende Gesprächsinhalte und fehlende Interaktion* (n = 6): Das Gespräch sei inhaltlich nur oberflächlich verlaufen aufgrund fehlender thematischer Vertrautheit, geringer Motivation oder mangelnden Interesses.
3. *Fehlender Erkenntnisgewinn* (n = 2): In dem Gespräch haben die Studierenden nichts gelernt.
4. *Störende Faktoren* (n = 1): Der Lärm im Zimmer hat ein intensives Gespräch verhindert.

6. Diskussion

Ziel des Beitrags war es, Potenziale und Grenzen eines LLM-basierten sozialen Netzwerkinderviews im Vergleich zu Peer-Gesprächen zwischen Studierenden zu untersuchen. Im Fokus standen vier Themenbereiche: Nachfrageverhalten, Antwortlänge, Deckungsgleichheit zwischen Netzwerkkarten und Interviews bzw. Gesprächen sowie das subjektive Erleben der Erhebungsformate.

Die Ergebnisse lassen sich wie folgt zusammenfassen: Erstens stellt der Chatbot, erwartungskonform, signifikant mehr Nachfragen als Studierende in Peer-Gesprächen. Zweitens führt diese erhöhte Nachfragefrequenz jedoch nicht zu längeren Antworten; bezogen auf den Zeichenumfang werden im LLM-Interview keine umfangreicheren Informationen generiert als im Gespräch. Drittens zeigt sich hinsichtlich der Deckungsgleichheit, dass in beiden Formaten vergleichbare Anteile von Akteur:innen unbesprochen bleiben, obwohl sie in der Netzwerkkarte enthalten sind. Gleichzeitig werden in den Peer-Gesprächen signifikant häufiger zusätzliche Akteur:innen genannt, die nicht in der Netzwerkkarte aufgeführt waren, was auf einen zusätzlichen Informationsgewinn durch

die Gespräche hinweist. Viertens wird das LLM-Interview als ambivalent erlebt: einerseits strukturiert, elaboriert und reflexionsfördernd, andererseits mit repetitivem und zu hartnäckigem Frageverhalten und teilweise kognitiv überfordernd. Die Peer-Gespräche werden insbesondere für ihre offene Gesprächsatmosphäre und den wechselseitigen Erkenntnisgewinn geschätzt, werden jedoch teils als oberflächlich beschrieben und gehen, in Übereinstimmung mit früheren Befunden (Galle, Kreis, und Hiebler 2025), nur selten mit vertiefenden Nachfragen einher.

6.1 Methodologische Diskussion zum LLM-basierten sozialen Netzwerkinterview (LLM-Interview)

Die Ergebnisse lassen sich entlang der oben eingeführten Kriterien leitfadengestützter Interviews einordnen: Offenheit, Spezifität, Kontextualität und Relevanz (Helfferich 2014; Przyborski und Wohlrab-Sahr 2021).

Hinsichtlich der *Offenheit* zeigt sich, dass das LLM-Interview entlang der Netzwerkkarte strukturiert ist. Zwar ermöglicht dies fokussierte Beschreibungen von Lerngelegenheiten, zugleich begrenzen hohe Nachfragedichte und Wiederholungen die Offenheit des Gesprächs. Teilweise ähnelt der Interaktionsverlauf eher einem schematischen Abarbeiten als einem dialogischen Interview, ein Befund, der auch von Wuttke u. a. (2025) berichtet wird.

Die *Spezifität* des Chatbots zeigt sich in den immanenten Nachfragen (Przyborski und Wohlrab-Sahr 2021) zu Situationen, Beispielen und Interaktionen. Diese werden konsequent umgesetzt und tragen zur Präzisierung vager Aussagen bei. Gleichzeitig wird diese Stärke von Studierenden auch als Überforderung erlebt, wenn Nachfragen an Punkten persistieren, an denen aus ihrer Sicht keine weiteren Inhalte genannt werden können. Es zeigt sich, dass trotz der deutlich höheren Zahl spezifischer Nachfragen keine längeren Antworten resultieren als in den Peer-Gesprächen. Die Erwartung, dass eine höhere Nachfragefrequenz zu längeren Antworten führt, wird somit nicht bestätigt. Damit hängt der Informationsgewinn qualitativer Interviews nicht allein von der Häufigkeit der Nachfragen ab, sondern ebenso von deren situativer Passung, ihrem wahrgenommenen Mehrwert sowie ihrer Einbettung in den Gesprächsverlauf. Die subjektiven Rückmeldungen der Studierenden liefern hierfür einen möglichen Erklärungsansatz: Wiederholte oder als schematisch wahrgenommene Nachfragen könnten die Bereitschaft zu narrativen Ausführungen eher begrenzen als fördern.

Bezüglich der *Kontextualität* stößt das LLM-Interview an Grenzen. Zwar werden Settings, Lernanlässe und Bedeutungszuschreibungen systematisch erfragt, jedoch fehlen dem Chatbot weiterführende Kontextinformationen, um situative Bedeutungen interpretativ aufzugreifen. Dies bestätigt die grundsätzliche Kritik, dass LLMs über algorithmische Kohärenz verfügen, nicht jedoch über ein verstehendes, empathisches Eingehen auf subjektive Bedeutungen (Williams und Ingleby 2025; Zhang u. a. 2025).

Hinsichtlich der *Relevanz* ist positiv hervorzuheben, dass der Chatbot thematisch fokussiert bleibt und den Leitfaden konsequent einhält (Helfferich 2014; Kruse 2015). Gleichzeitig richtet sich die häufigste Kritik der Studierenden auf persistente und detailierte Nachfragen, denen aus ihrer Sicht kein erkennbarer Mehrwert zukommt. Auch hier zeigt sich: LLMs produzieren kohärente, aber nicht notwendigerweise inhaltlich bedeutsame Fragen (Wuttke u. a. 2025).

Zusammenfassend erfüllt das eingesetzte LLM-Interview zentrale formale Kriterien leitfadengestützter Interviews, insbesondere hinsichtlich der thematischen Fokussierung und der konsequenten Umsetzung immanenter Nachfragen. Die Ergebnisse zeigen jedoch zugleich, dass diese formale Gesprächsführung allein keinen höheren Informationsgewinn gewährleistet. Grenzen werden insbesondere dort sichtbar, wo Offenheit, Kontextualität und die adaptive Einbettung von Nachfragen in den Gesprächsverlauf erforderlich sind; Merkmale, die eng mit situativem Verstehen und Empathie verbunden sind.

6.2 LLM-Interviews und Mixed-Methods-Forschung

Über die Kriterien qualitativer Interviews hinaus lassen sich LLM-Interviews insbesondere im Kontext der Mixed-Methods-Forschung diskutieren. Ein erster Mehrwert liegt in Conversion-Mixed-Methods-Designs (Tashakkori, Johnson, und Teddlie 2021). Durch die standardisierte und skalierbare Datenerhebung können qualitative Daten ohne zusätzlichen personellen Aufwand in größerem Umfang erhoben werden. Diese Daten lassen sich kodieren, in quantitative Variablen überführen und anschliessend sowohl qualitativ als auch quantitativ analysieren (Galle und Iten 2027).

Ein zweiter Mehrwert betrifft die Integration qualitativer und quantitativer Daten bereits im Erhebungsprozess. Gerade in der sozialen Netzwerkforschung stellt die zeitgleiche Erhebung beider Datentypen eine methodische Herausforderung dar. Aufgrund organisatorischer Rahmenbedingungen erfolgen quantitative Netzwerkerhebungen und qualitative Interviews häufig zeitlich versetzt. Da soziale Netzwerkstrukturen dynamisch sind und sich innerhalb weniger Tage oder Wochen verändern können, besteht das Risiko, dass sich beide Datenerhebungen auf unterschiedliche relationale Strukturen beziehen (Fröhlich, Rehm und Rienties 2020). Das hier vorgestellte Erhebungsinstrument begegnet diesem Problem, indem qualitative und quantitative Daten unmittelbar miteinander verknüpft werden. Das LLM-Interview baut direkt auf der zuvor erstellten Netzwerkkarte auf, sodass sich die qualitativen Erläuterungen auf dieselben Akteur:innen und relationalen Strukturen beziehen, die zuvor quantitativ erfasst wurden. Dadurch erfolgt die Integration beider Datentypen bereits während der Datenerhebung und nicht erst im Rahmen der Datenanalyse.

6.3 Potenziale für die Weiterentwicklung des Chatbots

Ein zentrales Ergebnis besteht darin, dass mehr Nachfragen nicht automatisch zu mehr inhaltlicher Tiefe führen. Die signifikant höhere Nachfragesfrequenz im LLM-Interview schlägt sich nicht in längeren Antworten nieder. Dies verdeutlicht erstens, dass die Anzahl von Nachfragen kein hinreichender Indikator für inhaltliche Vertiefung ist. Zweitens legt es nahe, dass die Interaktion mit dem Chatbot von Studierenden teilweise als stark strukturierte Befragung wahrgenommen wird. Wiederholte, ähnlich formulierte Nachfragen scheinen eher zu knappen Antworten zu führen als zu narrativen Ausführungen. In Folgestudien sollte getestet werden, wie die Antwortlänge ausfällt, wenn keine schriftlichen, sondern mündlichen Antworten gegeben werden. Unter Voraussetzung eines ruhigen Umfelds könnten möglicherweise deutlich mehr Informationen innert kürzester Zeit von den Studierenden erfasst werden.

Mit Blick auf die Netzwerkkarten zeigt sich ein differenziertes Bild. Positiv ist hervorzuheben, dass der Grossteil der Netzwerkkarten in beiden Gruppen vollständig erhoben wurde. Problematisch ist jedoch, dass auch im LLM-Interview relevante Akteur:innen unbesprochen bleiben, obwohl der Chatbot explizit angewiesen war, alle Personen zu thematisieren. Ein möglicher Erklärungsansatz liegt in der zeitlichen Begrenzung der Interviews. Zudem zeigt der Vergleich mit der DiaMaNt-Studie (Galle, Kreis und Hiebler 2025), dass die Deckungsgleichheit zwischen Netzwerkkarte und Gespräch dort höher ausfiel. Möglicherweise liegt dies an der unterschiedlichen Erzählaufforderung: Während in der DiaMaNt-Studie konkrete Lernsituationen im Zentrum standen, fokussierte die vorliegende Studie stärker den abstrakteren Gegenstand «lernzentrierte Kooperation». Offenbar lädt das Erzählen über konkrete Situationen eher dazu ein, alle relevanten Akteur:innen zu thematisieren.

Darüber hinaus werden in den Peer-Gesprächen signifikant häufiger zusätzliche Akteur:innen genannt als im LLM-Interview. Hier bleibt das LLM-Interview hinter der menschlichen Interaktion zurück. Es scheint, dass sich der Chatbot primär zur Strukturierung und Spezifizierung vorhandener Informationen eignet, weniger zur Erfassung neuer Akteur:innen. Eine Ursache könnte im Frageverhalten des Chatbots liegen. Der Chatbot konnte nicht «abschätzen», wann ein Inhalt ausgeschöpft ist. Eine mögliche Weiterentwicklung bestünde darin, Promptlogiken zu implementieren, die den inhaltlichen Ausschöpfungsgrad erkennen und gezielt zwischen vertiefenden und öffnenden Fragen wechseln, um den Interviewfluss zu verbessern.

7. Limitationen

Erstens folgt diese Studie einem explorativen Design und ist keine Experimentalstudie. Trotz einzelner Standardisierungen (z. B. einheitliche Einführung und Erklärungen mittels Videos) lassen sich bestimmte Störvariablen (z. B. intrapersonale Merkmale der Studierenden) nicht vollständig kontrollieren. *Zweitens* erfolgte die zufällige Gruppeneinteilung

auf Ebene der Mentoratsgruppen und nicht auf Individualebene, wodurch gruppenspezifische Einflüsse (z. B. soziales Klima in der Gruppe) nicht vollständig ausgeschlossen werden können. *Drittens* ist der Vergleichsmassstab kritisch zu beurteilen: Peer-Gespräche stellen keinen methodischen «Goldstandard» qualitativer Interviews dar, sondern ein Gespräch zwischen Personen mit wenig forschungsmethodischer Expertise. Ein Vergleich mit von Forschenden durchgeführten Netzwerkinderviews wäre aussagekräftiger. *Viertens* ist der Rücklauf der Audiodaten geringer als jener der LLM-Interviews, was die statistische Power der Gruppenvergleiche einschränkt. *Fünftens* fokussiert die vorliegende Analyse primär formale Indikatoren und subjektive Einschätzungen der Studierenden; die inhaltliche Qualität der Aussagen der Studierenden wurde bislang nicht systematisch untersucht. *Sechstens* ist die Studie im Feld der sozialen Netzwerkforschung verortet. Somit ist die Übertragung der Ergebnisse auf andere Kontexte nur begrenzt möglich.

8. Fazit

Die Studie zeigt, dass LLM-basierte soziale Netzwerkinderviews neue Möglichkeiten für eine skalierbare, standardisierte und zugleich fallnahe Datenerhebung eröffnen. Ihre Stärke liegt insbesondere in der systematischen Gesprächsführung, der hohen Nachfragesfrequenz sowie der Möglichkeit, qualitative Daten in grösseren Stichproben standardisiert zu erheben und unmittelbar für Mixed-Methods-Analysen verfügbar zu machen.

Gleichzeitig werden deutliche Grenzen sichtbar. LLM-Interviews in dieser Studie erzeugen zwar kohärente und formal anschlussfähige Interaktionen, erreichen jedoch nur begrenzt jene Offenheit, Kontextualität und subjektive Relevanz, die qualitative Interviews häufig durch situatives Verstehen, Empathie und Beziehungsarbeit entfalten. Mehr Nachfragen führen dabei nicht automatisch zu grösserer inhaltlicher Tiefe oder zu zusätzlichen narrativen Öffnungen.

LLM-Interviews erscheinen daher weniger als Ersatz qualitativer Interviews, sondern vielmehr als eigenständiges Erhebungsformat in Mixed-Methods-Designs, bei denen eine zeitgleiche Erhebung qualitativer und quantitativer Daten bedeutsam ist; zum Beispiel in der sozialen Netzwerkforschung.

Literatur

- Bauer, Petra, und Melanie Fabel-Lamla. 2020. «(Multi-)Professionelle Kooperation in der Lehrerinnen- und Lehrerbildung». In *Handbuch Lehrerinnen- und Lehrerbildung*, herausgegeben von Colin Cramer, Johannes König, Martin Rothland und Sigrid Blömeke, 91–97. Bad Heilbrunn: Klinkhardt. <https://doi.org/10.35468/hblb2020-010>.
- Bellotti, Elisa. 2015. *Qualitative networks. Mixed Methods in sociological research*. London: Routledge.

- Brennan, Robert L., und Dale J. Prediger. 1981. «Coefficient Kappa: Some uses, misuses, and alternatives». *Educational and Psychological Measurement* 41: 687–99.
- Burleigh, Cheryl, und Andrea M. Wilson. 2024. «Generative AI: Is authentic qualitative research data collection possible?». *Journal of Educational Technology Systems* 53 (2): 89–115. <https://doi.org/10.1177/00472395241270278>.
- Chopra, Felix, und Ingar Haaland. 2023. «Conducting qualitative interviews with AI». Preprint, Rochester, NY. <https://doi.org/10.2139/ssrn.4583756>.
- Creswell, John W., und Vicki L. Plano Clark. 2018. *Designing and conducting mixed methods research*. 3. Auflage. Thousand Oaks: SAGE.
- Dröge, Kai. 2023. *noScribe. Cutting edge AI technology for automated audio transcription*. Luzern.
- Engeström, Yrjö. 1999. «Activity theory and individual and social transformation». In *Perspectives on activity theory*, herausgegeben von Yrjö Engeström, Reijo Miettinen und Raija-Leena Punamäki, 19–38. Cambridge; England: Cambridge University Press. <https://doi.org/10.1017/CBO9780511812774.003>.
- Fabel-Lamla, Melanie, und Cornelia Gräsel. 2020. «Professionelle Kooperation in der Schule». In *Handbuch Schulforschung*, herausgegeben von Tina Hascher, Till-Sebastian Idel und Werner Helsper, 1–21. Wiesbaden: Springer Fachmedien. https://doi.org/10.1007/978-3-658-24734-8_57-1.
- Fröhlich, Dominik E. 2023. «Mixed methods and social network analysis». In *International Encyclopedia of Education (Fourth Edition)*, herausgegeben von Robert J. Tierney, Fazal Rizvi und Kadriye Ercikan, 685–92. Oxford: Elsevier. <https://doi.org/10.1016/B978-0-12-818630-5.11059-0>.
- Fröhlich, Dominik E., M. Rehm und B. C. Rienties, Hrsg. 2020. *Mixed methods social network analysis. Theories and methodologies in learning and education*. Abingdon, NY: Routledge.
- Galle, Marco, und Glena Iten. 2027. *Kodierungen analysieren und visualisieren. Mixed-Methods-Datenanalyse mit MAXQDA und R*. Springer.
- Galle, Marco, Annelies Kreis, und Sonja Hiebler. 2022. *Social network research with PowerPoint. Key moments in student teachers' learning during an internship*. Lucerne: University of Teacher Education Lucerne. <https://doi.org/10.13140/RG.2.2.30664.60164>.
- Galle, Marco, Annelies Kreis, und Sonja Hiebler. 2025. «Social networks and key situations of student-teacher learning regarding teaching». Konferenzbeitrag [single paper]. EARLI Conference. August 2025.
- Galle, Marco, Mirco Strässle, und Matthias Loser. 2025. *Social network interviews with large language models (artificial intelligence). A study on student-centred collaboration in the social network «teacher education». A mixed methods data collection approach on the collaborative activities related to the learning processes of students*. Lucerne: University of Teacher Education Lucerne. <https://doi.org/10.13140/RG.2.2.22546.36800>.
- Galle, Marco. 2026. «Collaboration within school social networks: a quantitative analysis of syllabi in Swiss teacher education». *Journal of Curriculum Studies*. <https://dx.doi.org/10.1080/00220272.2026.2634700>.

- Grosche, Michael, und Elisabeth Moser Opitz. 2023. «Kooperation von Lehrkräften zur Umsetzung von inklusivem Unterricht – notwendige Bedingung, zu einfach gedacht oder überbewerteter Faktor?» *Unterrichtswissenschaft* 51 (2): 245–63. <https://doi.org/10.1007/s42010-023-00172-3>.
- Gruber, Hans, Markus Hirschmann, und Monika Rehl. 2018. «Bildungsbezogene Netzwerkforschung». In *Handbuch Bildungsforschung*, 4. Auflage, herausgegeben von Rudolf Tippelt und Bernhard Schmidt-Hertha, 1339–56. Springer Reference Sozialwissenschaften. Wiesbaden: Springer Fachmedien. https://doi.org/10.1007/978-3-531-19981-8_60.
- Han, Songyue, Mingyu Wang, Jialong Zhang, Dongdong Li, und Junhong Duan. 2024. «A review of Large Language Models: Fundamental architectures, key technological evolutions, interdisciplinary technologies integration, optimization and compression techniques, applications, and challenges». *Electronics* 13 (24): 5040. <https://doi.org/10.3390/electronics13245040>.
- Helfferrich, Cornelia. 2011. *Die Qualität qualitativer Daten. Manual für die Durchführung qualitativer Interviews*. 4. Wiesbaden: Verlag für Sozialwissenschaften.
- Helfferrich, Cornelia. 2014. «Leitfaden- und Experteninterviews». In *Handbuch Methoden der empirischen Sozialforschung*, herausgegeben von Nina Baur und Jörg Blasius, 559–75. Wiesbaden: Springer.
- Hollstein, Betina, Tom Töpfer, und Jürgen Pfeffer. 2020. «Collecting egocentric network data with visual tools: A comparative study». *Cambridge University Press* 8 (2): 223–50. <https://doi.org/10.1017/nws.2020.4>.
- Keller-Schneider, Manuela, und Uwe Hericks. 2020. «Berufseinstieg von Lehrerinnen und Lehrern». In *Handbuch Schulforschung*, herausgegeben von Tina Hascher, Till-Sebastian Idel und Werner Helsper, 1–20. Wiesbaden: Springer Fachmedien. https://doi.org/10.1007/978-3-658-24734-8_59-1.
- Kreis, Annelies, und Esther Brunner. 2022. «Berufspraktische Lehrpersonenbildung als Tätigkeit in sozialen Netzwerken: theoretischer Rahmen und methodische Konzeption für eine interdisziplinäre Analyse aus allgemein- und mathematikdidaktischer Perspektive». In *Schul- und Berufspraktische Studien und die Fachdidaktiken. Verhältnisbestimmungen – Methoden – Empirie*, herausgegeben von Tobias Leonhard, Thomas Royer, Matthias Schierz, Christine Streit und Esther Wiesner, 7:179–201. Münster: Waxmann.
- Kruse, Jan. 2015. *Qualitative Interviewforschung*. 2. Auflage. Weinheim: Beltz.
- Kuckartz, Udo, und Stefan Rädiker. 2024. *Qualitative Inhaltsanalyse. Methoden, Praxis, Umsetzung mit Software und künstlicher Intelligenz*. 6., überarbeitete und erweiterte Auflage. Beltz.
- Liu, Zhe, Jiamin Dai, Cristina Conati und Joanna McGrenere. 2025. «Envisioning AI support during semi-structured interviews across the expertise spectrum». *Proc. ACM Hum.-Comput. Interact.* 9 (2): 1–29. <https://doi.org/10.1145/3710909>.
- Loosen, Wiebke. 2016. «Das Leitfadeninterview – eine unterschätzte Methode». In *Handbuch nicht standardisierte Methoden in der Kommunikationswissenschaft*, herausgegeben von Stefanie Auerbeck-Lietz und Michael Meyen, 139–55. Wiesbaden: Springer Fachmedien.
- McCormick, R., A. Fox, P. Carmichael, und R. Procter. 2011. *Researching and understanding educational networks*. New York: Routledge.

- Przyborski, Aglaja, und Monika Wohlrab-Sahr. 2021. *Qualitative Sozialforschung: Ein Arbeitsbuch*. De Gruyter Oldenbourg.
- Resnick, Lauren B., J. Levine, und S. Teasley, Hrsg. 1991. *Perspectives on socially shared cognition*. Hyattsville, MD: American Psychological Association.
- Ronfeldt, Matthew. 2015. «Field placement schools and instructional effectiveness». *Journal of Teacher Education* 66 (4): 304–20. <https://doi.org/10.1177/0022487115592463>.
- Saka, Adewale Owodunni. 2021. «Can teacher collaboration improve students' academic achievement in junior secondary mathematics?». *Asian Journal of University Education* 17 (1): 33–64. <https://doi.org/10.24191/ajue.v17i1.8727>.
- Schönhuth, Michael, und Markus Gamper. 2013. «Visuelle Netzwerkforschung. Eine thematische Annäherung». In *Visuelle Netzwerkforschung. Qualitative, quantitative und partizipative Zugänge*, herausgegeben von Michael Schönhuth, Markus Gamper, Michael Kronenwett und Martin Stark, 9–32. Bielefeld: Transcript Verlag.
- Stegbauer, Christian. 2025. *The Twelve Basic Assumptions of Network Research*. Wiesbaden: Springer. <https://doi.org/10.1007/978-3-658-48025-7>.
- Strässle, Mirco. 2025. *smartive/phlu-network-ai-study*. TypeScript. 5. Mai 2025; Smartive, released Oktober 23. <https://github.com/smartive/phlu-network-ai-study>.
- Tashakkori, Abbas, R. Burke Johnson und Charles Teddlie. 2021. *Foundations of Mixed Methods Research: Integrating quantitative and qualitative approaches in the social and behavioral sciences*. 2. Edition. Los Angeles: SAGE.
- VERBI Software. 2026. *MAXQDA* (Version 24) [Computer software]. <https://www.maxqda.com>.
- Williams, Ryan Thomas, und Ewan Ingleby. 2025. «The online survey in qualitative research: can AI act as a probing tool?». *Frontiers in Research Metrics and Analytics* 10: 1–6. <https://doi.org/10.3389/frma.2025.1519008>.
- Winkler, Anja, und Daniela Freisler-Mühlemann. 2023. «Einstieg in den Lehrberuf – Formen des Umgangs von Lehrpersonen mit professionellen Anforderungen». *Lehrerbildung auf dem Prüfstand* 37 (2): 5–24.
- Wuttke, Alexander, Matthias Aßenmacher, Christopher Klamm, Max M. Lang, Quirin Würschinger, und Frauke Kreuter. 2025. «AI Conversational Interviewing: Transforming Surveys with LLMs as Adaptive Interviewers». In *Proceedings of the 9th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature (LaTeCH-CLfL 2025)*, herausgegeben von Anna Kazantseva, Stan Szpakowicz, Stefania Degaetano-Ortlieb, Yuri Bizzoni und Janis Pagel, 179–204. Albuquerque, New Mexico: Association for Computational Linguistics. <https://doi.org/10.18653/v1/2025.latechclfl-1.17>.
- York-Barr, Jennifer, Gail Ghere, und Jennifer Sommersness. 2007. «Collaborative teaching to increase ELL student learning: A three-year urban elementary case study». *Journal of Education for Students Placed at Risk* 12 (3): 301–35. <https://doi.org/10.1080/10824660701601290>.
- Zhang, He, Chuhao Wu, Jingyi Xie, Yao Lyu, Jie Cai, und John M. Carroll. 2025. «Harnessing the power of AI in qualitative research: Exploring, using and redesigning ChatGPT». *Computers in Human Behavior: Artificial Humans* 4 (2025): 100144. <https://doi.org/10.1016/j.chbah.2025.100144>.

Danksagung

Wir danken den Mentorierenden der Pädagogischen Hochschule Luzern für ihre Unterstützung bei der Durchführung der Erhebung. Unser besonderer Dank gilt den Studierenden, die mit ihrer Teilnahme und ihren reflektierten Rückmeldungen wesentlich zum Gelingen der Studie beigetragen haben. Die Studie wurde durch den Schweizerischen Nationalfonds (SNF) gefördert, dem wir für die finanzielle Unterstützung danken.