

Qualitative Content Analysis With ChatGPT: Pitfalls, Rough Approximations and Gross Errors. A Field Report

Philipp Mayring

Key words:

ChatGPT; artificial intelligence; qualitative content analysis; theme analysis; qualitative text analysis; grounded-theory-methodology; psychoanalytical text interpretation

Abstract: In this article I describe a series of test runs to examine the contribution that the AI-based program ChatGPT in both Versions 3.5. and 4 can make to a qualitative content analysis of interview texts. A short sample text with my sample solution is presented for this purpose. Rough inputs for a rather naive use ("Conduct a qualitative content analysis!") as well as differentiated specifications with questions and more precise coding instructions (prompts) led in both versions at most to rough approximations of the sample solution with a large number of gross errors. The program did not react or reacted incorrectly to different content analysis concepts (BRAUN & CLARKE, 2006; KUCKARTZ, 2014; MAYRING, 2022a; SCHREIER, 2012), did not recognize hidden text content, and failed to check for coding agreement. The results of the software, no matter what specifications were made, mostly pointed in the direction of a rough, superficial summary in the sense of a list of topics and thus appear to be less suitable for the qualitative content analysis methods I developed (MAYRING, 2022a, 2022b).

Table of Contents

- [1. Introduction](#)
- [2. The Sample Text](#)
- [3. Qualitative Content Analysis With ChatGPT](#)
- [4. Working With More Precise Prompts](#)
- [5. Coder Agreements](#)
- [6. Testing With ChatGPT 4 and With Larger Amount of Text Material](#)
- [7. Conclusion](#)

[References](#)

[Author](#)

[Citation](#)

1. Introduction

I recently experienced the following during a university event on artificial intelligence (AI) in research and teaching: A lecturer reported that he liked to advise the students for Bachelor's and Master's theses, if they had conducted interviews for data collection, to first enter the material into ChatGPT and ask the AI to perform a qualitative content analysis according to MAYRING (2022b). I was perplexed. For over 40 years, my team and I have been refining qualitative content analysis techniques with step models and procedural rules. Interactive software ([QCAmap](#)) has been developed to support this, and now ChatGPT is supposed to do the whole thing automatically in a matter of seconds without textbook study, without consultation, without an introductory workshop? [1]

To test the performance of ChatGPT, I present here a sample text and a sample solution of a qualitative content analysis developed in our working group (Section

2). In a whole series of tests, I requested a content analysis of ChatGPT (Section 3), also with precise prompt input (Section 4) as well as larger amounts of material (Section 6), and compared it with our sample solution. The problem of a lack of coding agreement on re-input is discussed in Section 5 and a rather sobering conclusion is drawn at the end (Section 7). [2]

2. The Sample Text

I immediately sat down at my desk at home and tried it out by using a text that we often use as illustrative material in our workshops. The excerpt comes from an interview with a former teacher in East Germany who became unemployed after reunification (MAYRING, KÖNIG, BIRK & HURST, 2000).

"P¹: The first time I was dismissed as a vocational schoolteacher was because, among many other subjects, I taught a subject, GDR-civics,² which is no longer included in the timetable of the state of Saxony. After five months of unemployment, I then worked for 10 months in a private educational institution for the retraining of unemployed workers and taught subjects such as general economics, business administration, commercial arithmetic, bookkeeping and everything else that came up. But these companies are no longer being allocated jobs by the employment office, so I've been unemployed for the second time since April.

I: Is this unemployment situation stressful for you at the moment?

P: Yes, it is stressful, stressful in the sense that you compare yourself with other colleagues who are still working and you yourself know what you are worth and what you can and can't do. And I believe that many dismissals into unemployment have simply been made arbitrarily and that all justice has been lost. I have already tried three times to get my rights in the labor court and I hope that I will succeed at the fourth attempt.

I: Does unemployment also have a positive side for you?

P: Actually, I would say no, because I really, really enjoyed being a teacher, really, really enjoyed teaching and just tidying the apartment, tidying the garden or shopping, I don't feel fulfilled overall.

I: Do you think you can cope with the stress?

P: Yes.

I: What did you do, you've already mentioned it, with the complaint to the labor court, did you do anything else about the situation?

P: Well, I made use of all the opportunities offered to me through the Education and Science Trade Union, and the complicated thing is that in the process after reunification. Many things, such as legal awareness, work with the works council and many other things, with which the West German citizen has grown up in 40 years of the Federal Republic, through the parental home, through school, through education, through the environment, all of which the East GDR citizen, ex-GDR citizen, had to learn in the last three or four years in addition to his daily work, in addition to the

1 P stands for the interviewee, I for the interviewer. The interview is from the cited project, but not printed in MAYRING et al. (2000), and has been translated into English for this article.

2 GDR = German Democratic Republic

course of his life. So, I believe that I used the opportunities that were available to me to further my education, including at the Federal Agency for Civic Education in M, for a one-week further education course, and I also made use of all other opportunities for literature etc., etc.

I: Would you draw a positive balance for your overall psychological situation?

P: Yes, I would say yes.

I: And the only stresses that are there are due to going to the job center?

P: The trips to the job center and simply because I don't yet know how the labor court proceedings will be decided. Because I also don't know how long my wife will still have a job and if I were to stop receiving unemployment benefit and my wife were unemployed, then it would be a real burden.

I: So, at the moment it's primarily the uncertainty of the situation.

P: Yes." [3]

I would first like to present my analysis. I refer to the concept of qualitative content analysis according to MAYRING (2022a, 2022b) and MAYRING and FENZL (2022). In contrast to other concepts (e.g., KUCKARTZ, 2014, SCHREIER, 2012, see Section 3), the aim is to assign categories as precisely as possible, which can be verified by coding agreements. To this end, clear procedural steps and process models were developed for each of the individual techniques proposed (summarization, inductive category formation, explication, deductive category assignment, mixed techniques), so that it makes sense to speak of a model solution. [4]

The question of the analysis was: What stress factors due to unemployment were reported? I then defined the technique to be used as inductive category formation. The category definition and the level of abstraction must be defined as content analysis rules. The category definition was "Subjectively negatively experienced aspects of changes in the life situation due to unemployment." The desired level of abstraction was a category formulation that was as concrete as possible in the text, but not so specific that the category could only be coded for the person in question. [5]

These rules result in six inductive categories:

- K1: Arbitrary, unfair dismissal
- K2: Loss of the beloved teaching profession
- K3: Household and garden not fulfilling
- K4: Unprepared for new conditions after the fall of communism
- K5: Uncertainty about the outcome of the lawsuit against dismissal
- K6: Financial uncertainty in the future [6]

"Going to the job center" was not categorized, as the coder agreed that it was only mentioned in response to an interview question (suggestive) and was immediately put into perspective. You "know what you are worth and what you

can and can't do" was not categorized as a violation of self-esteem, as this would have been too interpretative in comparison to the formulated level of abstraction and the category definition. [7]

3. Qualitative Content Analysis With ChatGPT

To test ChatGPT, I first used the free Version 3.5 and, following the "advice" of the colleague quoted in the introduction, I began quite openly: "Please do a qualitative content analysis according to Mayring with the following interview excerpt."³ This is, of course, a completely inadequate, naïve request, but one that is likely to occur frequently. We know this from the use of internet search engines, where users are usually satisfied with simple questions and superficial answers. In that respect, this is a pitfall of working with ChatGPT. The answer to the above question came very quickly (ChatGPT 3.5 from March 24, 2024):

"In order to carry out a qualitative content analysis according to Mayring, we will analyze the content of the interview excerpt and summarize it in categories. Here are the steps for the analysis:

Step 1: Transcribing the interview ...

Step 2: Categorization

Reasons for unemployment:

Civics no longer in the curriculum

Lack of assignments from the employment office

Stress caused by unemployment:

Comparison with other colleagues

Perception of injustice

Legal dispute in the labor court

Negative effects of unemployment:

Loss of the teaching profession

Lack of fulfillment through housework

Coping with the situation:

Belief in own ability to cope with stress

Use of trade union services and further training opportunities

Positive aspects of the situation:

Mental balance is positive overall

Uncertainty:

Uncertainty about the labor court proceedings

Uncertainty about the wife's future

3 The prompts and answers from ChatGPT were processed in German and translated into English

Step 3: Coding the material: The interview is coded according to the identified categories.

Step 4: Interpretation of the results: The frequency of the codes is analyzed to identify patterns and trends. The interpretation is made in the context of the research question or the aim of the study." [8]

The presentation of the method with the four steps listed is incorrect, has not been published anywhere in my work and is more similar to the procedure proposed by SCHREIER (2012) and criticized by me (MAYRING, 2020) as very time-consuming: First the category development, then the coding in a second material pass. The procedure developed by me and my team, on the other hand, consists of eight different techniques, from which a selection must be made according to the research question. If inductive category formation is chosen, a category definition and a level of abstraction must be determined after the content-analytical units (coding unit, context unit, recording unit) have been defined in a question-related and theory-based manner (see Section 2). Once categories have been developed, the possibility is offered to develop main categories based on theory, and not vice versa. The starting point is the material, which is then abstracted. [9]

ChatGPT seems to be based more on a summarizing qualitative content analysis, but does not adhere to the procedural steps for summaries proposed by me (MAYRING, 2022a, 2022b) (paraphrasing, reduction, generalization, further reduction[s] and generalization[s]). If one looks at the two negative categories, four of the six categories from the sample solution are roughly found in my model solution, a rather poor result, with three miscategorizations. One conclusion of my analysis (see Section 7) is already apparent: Rough approximations and gross errors. [10]

In another test, I asked the same question with a request for evaluation using qualitative content analysis according to SCHREIER (2012). In the answer (ChatGPT 3.5 from March 10, 2024), three steps were presented, namely open coding, axial coding and selective coding, which actually originate from grounded theory methodology (GTM, GLASER & STRAUSS, 1967; STRAUSS, 1987), but were incorrectly described and incorrectly performed by ChatGPT 3.5. According to GTM, the codes selected from open coding for axial coding should be those that appear to be useful for theory building. ChatGPT was unable to implement this. GTM was cited by SCHREIER (2012), but only as one possible step (alongside summarization, subsumption and contrasting) in the categorization of the material. Here, too, one could argue that the purpose of working with ChatGPT is not to ask unspecific global questions, but to approach them iteratively by specifying them in the question design (prompts) (LIEDER & SCHÄFFER, 2024). However, this would require a clear understanding of the method and the specific research question, which many users are unlikely to have. If you already had precise ideas at the beginning in order to be able to send meaningful prompts, you could carry out the analysis directly on the material without AI, which is another pitfall of using ChatGPT. An iterative process of

approaching an open question by refining the prompts is not really necessary when working with qualitative content analysis according to MAYRING (2022a), as clear questions and coding rules are available in advance. This may be different for more open, explorative approaches to qualitative research. [11]

The answer to the same question with the request for a qualitative content analysis according to KUCKARTZ (2014) was equally unsatisfactory. He had proposed three basic methods: Thematic, evaluative (deductive!) and type-building text analysis. The result presented by ChatGPT 3.5 (March 10, 2024) most closely resembled the first inductive technique, which should be explained. The individual categories and subordinate codes were similar to the solution according to SCHREIER (2012), as was the process of category formation and subsequent coding. [12]

The same experiment with the request for a thematic analysis according to BRAUN and CLARKE (2006, 2024) showed great similarities to the results determined for KUCKARTZ (2014), although the approach of BRAUN and CLARKE had previously been characterized in the ChatGPT response as aiming at hidden patterns and themes in the text. In contrast, BRAUN and CLARKE (2006, 2024) differentiated between inductive and deductive forms of thematic analysis. The aim of identifying hidden patterns corresponds most closely to the reflexive theme analysis described by them. The results presented by ChatGPT 3.5, however, did not contain any hidden patterns, but remained within the descriptive content. [13]

In further tests, I made a mistake and asked for a thematic analysis according to BROWN and HARRIS instead of BRAUN and CLARKE. BROWN and HARRIS have never dealt with text analysis, but are known for psychological life event and depression research (BROWN & HARRIS, 1978). Chat GPT nevertheless carried out an analysis in both Versions 3.5 and 4.0 and presented more or less meaningless results, for example, listing life events that were not even mentioned in the text. [14]

Further tests with ChatGPT showed that ironic remarks and indirect allusions were not recognized. ChatGPT also failed to recognize a text passage that contained a slip of the tongue and can be interpreted as an indicator of repression processes from a psychoanalytical perspective (see HAUBL and LOHL, 2020 on psychoanalytical text interpretation). More complex evaluation commands (e.g., "Please make me a list of the categories with the frequencies of the text passages that were found in each case") were not taken into account. When entering several interviews, I quickly reached a limit of around three transcript pages, which is an exclusion criterion for content analysis studies that often comprise several hundred pages. [15]

4. Working With More Precise Prompts

One reason for the unsatisfactory results is certainly that the analysis query was too global. Therefore, following the sample example, I started the next attempt with the more precise input:

"What stress factors due to unemployment does the person name? Please use the inductive category formation according to Mayring. Formulate the categories as specifically as possible in the text, but not so specifically that they would only appear for this one person. Only categorize if the negative factor is also subjectively significant." [16]

This corresponds to the rules formulated in my sample solution. The answer from ChatGPT 3.5 (from March 24, 2024) was significantly worse compared to the sample solution:

1. Comparison with other colleagues
2. Feeling of injustice during dismissals
3. Legal dispute before the labor court
4. Uncertainty about the outcome of the labor court proceedings
5. Uncertainty about the future of the wife [17]

The first mention is not an independent stress factor, but the injustice due to the comparison is stressful. The third mention is also not an independent burden, but an expression of uncertainty about the outcome of the proceedings. Three of the categories from the sample solution ("Loss of beloved teaching profession," "Household and garden not fulfilling," "Unprepared for new conditions after reunification"; see Section 2) were not recognized. [18]

The entries could be made more precise, with precise specifications and explanations of individual process steps and even sample analyses inserted in prompts. There are already freely available templates, so-called custom GPTs, such as a [Mayring Mentor](#), with over a thousand users⁴. I called up the template and saw the details of an author who described himself as an expert but, according to internet entries, works in the construction and real estate industry. The template contained serious errors: Only one of the techniques we developed was described, namely inductive category formation. This was mischaracterized, paraphrasing was required, which we only do for summaries, only one unit of analysis was specified (only the coding unit, not the context unit and recording unit, which are also necessary) and the step of main category formation was skipped. I suspect that ChatGPT had been used to create the template, and then we have gone round in circles. As someone who has spent decades teaching introductory courses in research methods, academic writing and colloquia on thesis supervision at university, I would like to shout out to the many users of this template: Texts that claim to be scientific must be sufficiently provided with sources that are verifiable ("Mayring Mentor" contains no references!). The

4 I would like to thank an anonymous reviewer for the information.

quality of the sources must be assured, for which the authorship should be checked (specialist studies, possibly doctorate, research experiences) and the quality of the publication medium should be considered (handbook or textbook article, specialist journal with review process, respected specialist publisher). Against this background, "Mayring Mentor" is not useful at all. The fact that the template reaches many users and is rated well cannot be a criterion. The fact that a lurid tabloid reaches many satisfied readers does not guarantee the truthfulness of the articles. [19]

5. Coder Agreements

I am not quite sure whether a re-run of the identical input in ChatGPT would rather measure intracoder agreement (the same person analyzes the material a second time in the sense of a reliability check) or intercoder agreement (a second person analyzes the material in the sense of an objectivity check). In any case, both are important quality criteria for qualitative content analysis (MAYRING, 2022a, 2022b). The result of the identical analysis task one day later produced the following ChatGPT expression (ChatGPT 3.5, March 25, 2024):

1. Comparison with colleagues on duty
2. Feeling of value and competence
3. Arbitrary dismissal
4. Legal dispute before the labor court
5. Dissatisfaction with current activities (not working as a teacher)
6. Lack of fulfillment through everyday tasks (household, gardening)
7. Use of support services (trade union, further training measures)
8. Challenges posed by the transition from the GDR to the Federal Republic in the work context
9. Uncertainty regarding the labor court proceedings
10. Concerns about the family's financial situation [20]

The result was even less useful, as not only stress factors were listed, as required in the prompt. Some things were now found that had been overlooked the day before. Overall, the coding agreement test must be considered completely failed. It is interesting to note that, even with Version 4, other solutions were always offered for identical entries during further checks of the coding match. I conclude from this that ChatGPT can only provide rough approximations at best. However, I also conclude from the above test run that ChatGPT does not always follow coding instructions and that this results in gross errors. Perhaps this was also due to the freely available Version 3.5 which I was using. [21]

6. Testing With ChatGPT 4 and With Larger Amount of Text Material

As mentioned above, ChatGPT showed a limit to the amount of material that can be processed (4096 tokens, about three pages of transcript, as ChatGPT reported to me). With the paid Version 4, approx. 500 pages of transcript can be processed. On request, however, it is pointed out that the most effective and accurate analysis is only guaranteed with smaller quantities, and it is recommended that the material be divided into portions. However, this is not very helpful for comprehensive analyses. Qualitative content analysis is often used to cross-analyze more than a thousand pages of material. [22]

I attempted a content analysis with about 10 pages of interview transcripts, excerpts from six interviews with unemployed people about their stresses. Here I asked ChatGPT 4 for a list of stress factors across the six cases, combined with a request to list the frequency of the categories found, a standard procedure of inductive category formation. The response is as follows (ChatGPT 4, March 18, 2024):

- "1. Comparison with colleagues who are still in service - 1 time
2. Arbitrary dismissals - 1 time
3. Legal disputes - 1 time
4. Lack of recognition of professional skills - 1 time
5. Difficulties in finding a job - 1 time
6. Uncertainty about the future - 1 time
7. Financial insecurity - 1 time
8. Loss of professional self-image - 1 time
9. Lack of career prospects - 1 time
10. Social isolation - 1 time
11. Adaptation to new professional demands - 1 time
12. Lack of professional satisfaction - 1 time
13. Inactivity and boredom - 1 time
14. Pressure to reorient oneself professionally - 1 time
15. Family stress due to unemployment - 1 time
16. Difficulties in dealing with the status of unemployment - 1 time
17. Experience of injustice and devaluation of previous jobs - 1 time
18. Need for further training without concrete job prospects - 1 time
19. Stress due to changed life routines - 1 time
20. Loss of professional network - 1 time" [23]

A list of 20 inductive categories was provided, but contrary to the instructions, each was coded only once. This was commented on as follows: "Each of these points was counted as once per relevant mention, which shows the variety of stress factors caused by unemployment among the unemployed teachers

interviewed" (ChatGPT 4, March 18, 2024). This evaluation does not match the sample solution above. Five of the 20 stress factors mentioned (16 categories were found there, see Section 2) were similar, but were mentioned several times in the interviews in the sample solution. For example, the text passage "just tidying up the apartment or the garden doesn't fill me up" was not found in Case 1, in Case 2 it was coded as "inactivity and boredom" (Case 2: "I could go into the garden and weed the garden now, but inside I can't do it"). Four of the six interviewees reported that they considered their dismissal to be unfair, which ChatGPT 4 did not recognize. Many text passages were not evaluated as stress factors, even though they clearly are, as in Case 1: "I: Does unemployment also have positive aspects for you? P: Actually, I would say no, because I really enjoyed being a teacher, I really, really enjoyed teaching." In the sample solution, this was coded with the category: "Loss of the beloved teaching profession." On the other hand, categories were named to which no text passages can be clearly assigned. Ultimately, the categorization of ChatGPT 4 is of little use. [24]

7. Conclusion

ChatGPT has produced results in both Versions 3.5 and 4, some of which can be classified as rough approximations, but some of which can also be classified as gross errors. The closest approximation to the sample solution (with "natural intelligence") was achieved in the ChatGPT 3.5 version with short text and simple questions. The fact that these are only approximations is also shown by the fact that a reanalysis with identical questions in ChatGPT produced different results in both versions. [25]

It has to be said that the possibilities of working with ChatGPT are far from exhausted. Above all, the narrowing down with increasingly specific questions and the precise prompt formulation were not implemented to any great extent. However, such an approach requires considerable expertise and the most precise ideas on the part of the users, which should be rare as a rule. In this respect, one can speak of pitfalls in the use of ChatGPT. An iterative approach to the results also does not correspond to the research style of the qualitative content analysis developed by our working group. If such a tentative procedure were used, I would not speak of a qualitative content analysis. [26]

ChatGPT 4 also failed with the more complex question (inductive categories on stress factors due to unemployment across several interviews). Important text passages were skipped, categorizations were imprecise, similarities in the stress factors between the interviews were not seen (despite requests). More specific coding instructions were not followed, various approaches to text analysis were not applied correctly. ChatGPT also performed incorrect instructions (see above "Please analyze with Brown and Harris thematic analysis") and presented results. [27]

The problematic data protection situation should also be pointed out. On the one hand, sending an interview transcript to ChatGPT actually requires the explicit, written consent of the interviewee. The fact that an interview from 30 years ago

was used in the case mentioned in Section 2, that the interviewee had given his consent to the anonymized use of the interview at the time and that he has probably died in the meantime, does not hide the fact that the use of interview data in ChatGPT in compliance with the [European Union General Data Protection Regulation](#) (GDPR) is extremely questionable. Where ChatGPT provides information beyond the transcript ("Which steps of the content analysis according to Mayring?"), there was also the problem that the answers did not contain any literature references and were therefore not scientifically usable. [28]

What can ChatGPT do? After the various attempts, it seems to me that rough content categorizations of text passages are reasonably successful and that an overview of topics in the material can be given. This may be helpful if one chooses an evaluation technique that does not presuppose any specific restrictions and tends to remain descriptive. This seems to me to be the case with the thematic qualitative text analysis according to KUCKARTZ (2014) or the thematic analysis according to BRAUN and CLARKE (2006). If the relevant software developers have now built AI modules based on ChatGPT in ATLAS.ti and MAXQDA, for example, they are only supporting a certain type of text analysis. The integration of process models and categorization rules is less intended there. This is also the reason why a separate, freely accessible software package (QCAmapp) was developed for the qualitative content analysis developed by me and my team, in which the process steps are interactively anchored in a binding manner. [29]

ChatGPT is therefore less suitable when

- more specific techniques with more precise evaluation instructions are to be used, such as in qualitative content analysis according to MAYRING (2022a);
- the text analysis is to be more in-depth, also taking theoretical considerations into account, such as in reflexive theme analysis (BRAUN & CLARKE 2024), GTM or psychoanalytical text interpretation (see MAYRING, 2019 for the different results of different text analysis approaches, shown using an example text). However, the structured, theory-based, rule-guided text analysis procedure seems central to me, as it meets high scientific standards (e.g., coding agreement testing) and can be meaningfully integrated into mixed-method designs, which are becoming increasingly important today. [30]

In the university course I mentioned in Section 1, the argument was put forward that you could have ChatGPT analyze the text first and then rework it. However, this doesn't seem to make much sense to me, as you don't know exactly where the inaccuracies and errors are, so you have to work through all the material in any case. So, what's the point of the ChatGPT run? A rough approximation of the results can even have a distorting influence on the evaluation if the entire material has to be processed in each case. [31]

One could use ChatGPT in qualitative text analysis for a first, very open exploratory step. However, the use cases for such "new ground research" are

rather rare in social research; usually one approaches the material with a more specific research question and a differentiated interview guideline, and there, ChatGPT becomes less precise. It is also difficult to implement a theory-based analysis. [32]

A final argument could be the amount of material that can be processed with ChatGPT. On the one hand, this is only possible in the paid Version 4, and even then, there are limitations. Secondly, it cannot be a goal in social research to collect as much material as possible. For an election analysis in the Federal Republic of Germany, I will not survey 50 million people, but rather select a sample. Qualitative research is also about defining samples in such a way that they allow generalizations and permit an in-depth, thorough analysis. [33]

Finally, it should be emphasized that tools such as ChatGPT are constantly undergoing revisions, which is why this can only be an interim review. However, I think that such experience reports are important for assessing the potential role of AI in science. [34]

References

- Braun, Virginia & Clarke, Victoria (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3, 77-101.
- Braun, Virginia & Clarke, Victoria (2024). Thematic analysis. In [Norman K. Denzin](#), Yvonna S. Lincoln, Michael D. Giardina & Gaile S. Canella (Eds.), *The Sage handbook of qualitative research* (6st ed., pp.385-402). Los Angeles, CA: Sage.
- Brown, George W. & Harris, Tirril (1978). *Social origins of depression: A study of psychiatric disorder in women*. New York, NY: Free Press.
- Glaser, Barney G. & Strauss, Anselm L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago, IL: Aldine.
- Haubl, Rolf & Lohl, Jan (2020). Tiefenhermeneutik. In [Günter Mey](#) & [Katja Mruck](#) (Eds.), *Handbuch Qualitative Forschung in der Psychologie* (2nd edition, Vol. 2, pp.555-578). Wiesbaden: Springer VS.
- Kuckartz, Udo (2014). *Qualitative text analysis. A guide to methods, practice & using software*. London: Sage.
- Lieder, Fabio R. & Schäffer, Burkhard (2024). *Reconstructive social research prompting (RSRP). Distributed interpretation between AI and researchers in qualitative research*, <https://doi.org/10.31235/osf.io/d6e9m> [Date of Access: October 1, 2024].
- [Mayring, Philipp](#) (2019). Zentrale qualitative Auswertungsverfahren. In Marius Haring, Carsten Rohlf's & Michaela Gläser-Zikuda (Eds.), *Handbuch Schulpädagogik* (pp.859-868), Münster: Waxmann – utb.
- Mayring, Philipp (2020). Qualitative content analysis: Demarcation, varieties, developments. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 20(3), Art. 16, <http://dx.doi.org/10.17169/fqs-20.3.3343> [Date of Access: October 1, 2024].
- Mayring, Philipp (2022a). *Qualitative content analysis. A step-by-step guide*. London: Sage.
- Mayring, Philipp (2022b). *Qualitative Inhaltsanalyse. Grundlagen und Techniken* (13th ed.). Weinheim: Beltz.
- Mayring, Philipp & Fenzl, Thomas (2022). Qualitative Inhaltsanalyse. In Nina Baur & Jürgen Blasius (Eds.), *Handbuch Methoden der empirischen Sozialforschung* (3rd ed., pp.691-706). Wiesbaden: Springer VS.
- Mayring, Philipp; König, Joachim; Birk, Nils & Hurst, Alfred (2000). *Opfer der Einheit – Eine Studie zur Lehrerarbeitslosigkeit in den neuen Bundesländern*. Opladen: Leske + Budrich.

[Schreier, Margrit](#) (2012). *Qualitative content analysis in practice*. London: Sage

Strauss, Anselm (1987). *Qualitative analysis for social scientists*. Cambridge: Cambridge University Press.

Author

[Philipp MAYRING](#) is retired professor for research methods on psychology at the Institute for Psychology of Alps-Adria University Klagenfurt, Austria and chief of the Association for Supporting Qualitative Research ASQ, Klagenfurt, Austria. In his research, he focusses on qualitative text analysis, evaluation research, health and well-being.

Contact:

Philipp Mayring

Alps-Adria University Klagenfurt
Institute for Psychology
Universitaetsstrasse 65-67, 9020 Klagenfurt,
Austria

E-mail: philipp.mayring@aau.at

URL: <https://philipp.mayring.at/>

Citation

Mayring, Philipp (2025). Qualitative content analysis with ChatGPT: Pitfalls, rough approximations and gross errors. A field report [34 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 26(1), Art. 4, <https://doi.org/10.17169/fqs-26.1.4252>.