

Draft references for Baker, Ryan S.; Barany, Amanda (forthcoming) *Artificial Intelligence in Qualitative Research: New Possibilities*. Philadelphia, PA: Vassant Press.

Find the latest chapters online at

<https://learninganalytics.upenn.edu/book/ai-qualitative-research.html>

Copyright Ryan S. Baker and Amanda Barany, 2026, all rights reserved.

References

Version 1.06 July 15, 2026

AI Notkilleveryoneism Memes (2025) An engineer showed Gemini what another AI said about its code. <https://x.com/AISafetyMemes/status/2000620127054598508>

AI Weekly (2025) Palmer Lucky Explains His ChatGPT Hack to Get It to Do ANYTHING. https://www.youtube.com/shorts/qS4S_-p-zso

Amirova, A., Fteropoulli, T., Ahmed, N., Cowie, M. R., & Leibo, J. Z. (2024). Framework-based qualitative analysis of free responses of Large Language Models: Algorithmic fidelity. *Plos one*, 19(3), e0300024.

Andriushchenko, M., Croce, F., & Flammarion, N. Jailbreaking Leading Safety-Aligned LLMs with Simple Adaptive Attacks. (2023) In *The Thirteenth International Conference on Learning Representations*.

Anthropic. (2026, March 19). What 81,000 people want from AI. <https://www.anthropic.com/81k-interviews>

Argyle, L. P., Busby, E. C., Fulda, N., Gubler, J. R., Rytting, C., & Wingate, D. (2023). Out of one, many: Using language models to simulate human samples. *Political Analysis*, 31(3), 337–351

Ashwin, T. S., Shafi, S. D., & Ramkumar, R. (2023). Dlot: An open-source application to assist human observers. *International Conference on Computers in Education*..

Atil, B., Aykent, S., Chittams, A., Fu, L., Passonneau, R. J., Radcliffe, E., ... & Baldwin, B. (2024). Non-determinism of "deterministic" llm settings. *arXiv preprint arXiv:2408.04667*.

Bai, Y., Jones, A., Ndousse, K., Askell, A., Chen, A., DasSarma, N., ... & Kaplan, J. (2022). Training a helpful and harmless assistant with reinforcement learning from human feedback. arXiv preprint arXiv:2204.05862.

Baker, Ocumpaugh, & Andres (2020), "BROMP Quantitative Field Observations: A Review," in R. Feldman (Ed.), *Learning Science: Theory, Research, and Practice*.

Baker, R.S., Hawn, M.A. (2022) Algorithmic Bias in Education. *International Journal of Artificial Intelligence and Education*, 32, 1052-1092.

Baker, R. S., Hutt, S., Bosch, N., Ocumpaugh, J., Biswas, G., Paquette, L., ... & Munshi, A. (2024). Detector-driven classroom interviewing: Focusing qualitative researcher time by selecting cases in situ. *Educational technology research and development*, 72(5), 2841-2863.

Barab, S., & Squire, K. (2016). Design-based research: Putting a stake in the ground. In *Design-based research* (pp. 1-14). Psychology Press.

Barambones, J., Moral, C., de Antonio, A., Imbert, R., Martínez-Normand, L., & Villalba-Mora, E. (2024). ChatGPT for learning HCI techniques: A case study on interviews for personas. *IEEE Transactions on Learning Technologies*, 17, 1460-1475.

Barany, A., Nasiar, N., Porter, C., Zambrano, A. F., Andres, A. L., Bright, D., ... & Baker, R. S. (2024, July). ChatGPT for education research: Exploring the potential of large language models for qualitative codebook development. In *International conference on artificial intelligence in education* (pp. 134–149). Cham: Springer Nature Switzerland.

Barany, A., Shah, M., & Foster, A. (2021, January). Connecting curricular design and student identity change: an epistemic network analysis. In *International Conference on Quantitative Ethnography* (pp. 155-169). Cham: Springer International Publishing.

Baudrillard, J. (1981). *Simulacra and simulation*. University of Michigan press.

Bazeley, P. (2013). *Qualitative data analysis: Practical strategies*. Sage.

Bødker, S., Dindler, C., Iversen, O. S., & Smith, R. C. (2022). *Participatory design*. Berlin, Germany: Springer International Publishing.

Bødker, S., Kensing, F., & Simonsen, J. (2004). *Participatory IT design: Designing for business and workplace realities*. MIT Press.

Behrens, J. T. (1997). Principles and procedures of exploratory data analysis. *Psychological Methods*, 2(2), 131–160.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021, March). On the dangers of stochastic parrots: Can language models be too big? 🦜. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (pp. 610-623).

- Bennett, J. (2010). *Vibrant matter: A political ecology of things*. Duke University Press.
- Berger, R. (2015). Now I see it, now I don't: Researcher's position and reflexivity in qualitative research. *Qualitative research*, 15(2), 219-234.
- Beyer, H., Holtzblatt, K. (1997). *Contextual design: defining customer-centered systems*. Elsevier.
- Bijker, R., Merkouris, S. S., Dowling, N. A., & Rodda, S. N. (2024). ChatGPT for automated qualitative research: Content analysis. *Journal of Medical Internet Research*, 26, e59050.
- Blair, A. (2025) Gemini 3 is Evaluation-Paranoid and Contaminated.
https://www.lesswrong.com/posts/8uKQyjrAgCcWpfmcs/gemini-3-is-evaluation-paranoid-and-contaminated?utm_source=chatgpt.com
- Blandford, A., Furniss, D., & Makri, S. (2016). *Qualitative HCI research: Going behind the scenes*. Morgan & Claypool Publishers.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993-1022.
- Blumer, H. (1969). *Symbolic interactionism: Perspective and method*. University of California Press.
- Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2021). On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*.
- Booth, R (2026) From 'nerdy' Gemini to 'edgy' Grok: how developers are shaping AI behaviours. *The Guardian*, 4 February 2026,
<https://www.theguardian.com/technology/2026/feb/03/gemini-grok-chatgpt-claude-qwen-ai-chatbots-identity-crisis>
- Bosch, N., D'Mello, S.K., Ocumpaugh, J., Baker, R.S., Shute, V. (2016) Using video to automatically detect learner affect in computer-enabled classrooms. *ACM Transactions on Interactive Intelligent Systems*, 6 (2).
- Bostrom, N. (2015) *Superintelligence: Paths, Dangers, Strategies*. Oxford, UK: Oxford University Press.
- Bourdieu, P. (1984). *Distinction: A social critique of the judgement of taste* (R. Nice, Trans.). Harvard University Press.
- Boyatzis, R. E. (1998). *Transforming qualitative information: Thematic analysis and code development*. Sage.
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How people learn: Brain, mind, experience, and school*. National Academy Press.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2), 77-101.

Braun, V., & Clarke, V. (2013). *Successful qualitative research: A practical guide for beginners*. Sage.

Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative research in sport, exercise and health*, 11(4), 589-597.

Braun, V., & Clarke, V. (2021). One size fits all? What counts as quality practice in (reflexive) thematic analysis?. *Qualitative research in psychology*, 18(3), 328-352.

Braun, V., & Clarke, V. (2023). Toward good practice in thematic analysis: Avoiding common problems and being a knowing researcher. *International journal of transgender health*, 24(1), 1-6.

Braveman, P., & Gottlieb, L. (2014). The social determinants of health: It's time to consider the causes of the causes. *Public Health Reports*, 129(Suppl. 2), 19–31.
<https://doi.org/10.1177/00333549141291S206>

Bruner, J. (1990). *Acts of meaning*. Harvard University Press.

Bruner, J. (1991). The narrative construction of reality. *Critical Inquiry*, 18(1), 1–21.

Bryant, A., & Charmaz, K. (2011). Grounded theory. In *The SAGE handbook of innovation in social research methods* (pp. 205–227).

Bryant, A., & Charmaz, K. (Eds.). (2010). *The SAGE handbook of grounded theory*. Sage.

Buckingham Shum, S. (2024) Qreframer: a chatbot prompt that reveals your assumptions.
<https://oercommons.org/courseware/lesson/114039/overview>

Burawoy, M. (1998). The extended case method. *Sociological Theory*, 16(1), 4–33.

Cai, Z., Marquart, C., Eagan, B., Xiao, Y., & Williamson Shaffer, D. (2023). A Lightweight Interactive Regular Expression Generator for Qualitative Coding in Quantitative Ethnography. In *ICQE 2023 Proceedings* (pp. 455–469). Springer.

Carlini, N., Tramer, F., Wallace, E., Jagielski, M., Herbert-Voss, A., Lee, K., ... Erlingsson, Ú. (2021). Extracting training data from large language models. In *Proceedings of the 30th USENIX Security Symposium (USENIX Security '21)* (pp. 2633–2650).

Case, N. (2018). How to become a centaur. *Journal of Design and Science*, 3(5).

Charmaz, K. (1991). *Good days, bad days: The self in chronic illness and time*. Rutgers University Press.

Charmaz, K. (2014). *Constructing grounded theory* (2nd ed.). Sage.

Chatzichristos, G. (2025). Qualitative research in the era of AI: A return to positivism or a new paradigm? *International Journal of Qualitative Methods*.

Chen, H.-T. (1990). *Theory-driven evaluations*. Sage.

Chen, L., Zaharia, M., & Zou, J. (2024). How is ChatGPT's behavior changing over time? *Harvard Data Science Review*, 6(2).

Chiang, T. (2023, February 9). ChatGPT is a blurry JPEG of the web. *The New Yorker*.
<https://www.newyorker.com/tech/annals-of-technology/chatgpt-is-a-blurry-jpeg-of-the-web>

Choi, J., Ruis, A. R., Cai, Z., Eagan, B., & Shaffer, D. W. (2023). Does Active Learning Reduce Human Coding?: A Systematic Comparison of Neural Network with nCoder. In *ICQE 2022 Proceedings*. Springer.

Chopra, F., & Haaland, I. (2024). Conducting qualitative interviews with AI. *CESifo Working Papers*, 10666

Clandinin, D. J., & Connelly, F. M. (2000). *Narrative inquiry: Experience and story in qualitative research*. Jossey-Bass.

Clark, H. H., & Brown, K. (2006). Context and common ground. *Concise encyclopedia of philosophy of language and linguistics*, 85-87.

Clarke, A. E. (2005). *Situational analysis: Grounded theory after the postmodern turn*. Sage.

Clarke, A.E., Washburn, R., Friese, C., Clarke, A.E., & Washburn, R. (Eds.). (2015). *Situational Analysis in Practice: Mapping Research with Grounded Theory* (1st ed.). Routledge.
<https://doi.org/10.4324/9781315420134>

Cleo Nardo (2023) The Waluigi Effect (mega-post).
<https://www.lesswrong.com/posts/D7PumeYTDpFbTp3i7/the-waluigi-effect-mega-post>

Clifford, J. (1986). Introduction: Partial truths. In J. Clifford & G. E. Marcus (Eds.), *Writing culture: The poetics and politics of ethnography* (pp. 1–26). University of California Press.

Coffey, A., Beverley, H., & Paul, A. (1996). Qualitative data analysis: Technologies and representations. *Sociological research online*, 1(1), 80-91.

Collins, A. (1992). Toward a design science of education. In *New directions in educational technology* (pp. 15-22). Springer Berlin Heidelberg.

Conrad, P. (1985). The meaning of medications: Another look at compliance. *Social Science & Medicine*, 20(1), 29–37. [https://doi.org/10.1016/0277-9536\(85\)90308-9](https://doi.org/10.1016/0277-9536(85)90308-9)

Corbin, J., & Strauss, A. (2014). Basics of qualitative research: Techniques and procedures for developing grounded theory. Sage.

Cornish, F., Gillespie, A., & Zittoun, T. (2013). Collaborative analysis of qualitative data. In *The SAGE Handbook of Qualitative Data Analysis* (pp. 79–93).

Creswell, J. W., & Creswell, J. D. (2017). Research design: Qualitative, quantitative, and mixed methods approaches. Sage publications.

Creswell, J. W., & Poth, C. N. (2016). Qualitative inquiry and research design: Choosing among five approaches. Sage publications.

Crowston, K., Allen, E. E., & Heckman, R. (2012). Using natural language processing technology for qualitative data analysis. *International Journal of Social Research Methodology*, 15(6), 523-543.

Cuevas, A., Scurrall, J. V., Brown, E. M., Entenmann, J., & Daepf, M. I. (2025). Collecting qualitative data at scale with large language models: A case study. *Proceedings of the ACM on Human-Computer Interaction*, 9(2), 1-27.

Cummins, J., Elson, M., & Hussey, I. (2025). Cognitive dissonance in large language models is neither cognitive nor dissonant. *Proceedings of the National Academy of Sciences*, 122(35), e2517912122.

Darling-Hammond, L. (2006). Powerful teacher education: Lessons from exemplary programs. Jossey-Bass.

Davison, R. M., Chughtai, H., Nielsen, P., Marabelli, M., Iannacci, F., van Offenbeek, M., ... & Panteli, N. (2024). The ethics of using generative AI for qualitative data analysis. *Information Systems Journal*, 34(5), 1433-1439.

De Paoli, S., & Mathis, W. S. (2025). Reflections on inductive thematic saturation as a potential metric for measuring the validity of an inductive thematic analysis with LLMs. *Quality & Quantity*, 59(1), 683-709.

DeLanda, M. (2006). A new philosophy of society: Assemblage theory and social complexity. Continuum.

Delaney, J. (2025) Catching the Vibe of Vibe Coding. *Communications of the ACM*.
<https://cacm.acm.org/news/catching-the-vibe-of-vibe-coding/>

Deleuze, G., & Guattari, F. (1980). A thousand plateaus: Capitalism and schizophrenia (B. Massumi, Trans.). University of Minnesota Press.

Dell'Acqua, F., McFowland III, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., ... & Lakhani, K. R. (2023). Navigating the jagged technological frontier: Field experimental

evidence of the effects of AI on knowledge worker productivity and quality. Harvard business school technology & operations mgt. Unit working paper, (24-013).

Dengel, A., Gehrlein, R., Fernes, D., Görlich, S., Maurer, J., Pham, H. H., ... & Eisermann, N. D. G. (2023, October). Qualitative research methods for large language models: Conducting semi-structured interviews with ChatGPT and BARD on computer science education. In *Informatics* (Vol. 10, No. 4, p. 78).

Denzin, N. K., & Lincoln, Y. S. (2003). Introduction: The discipline and practice of qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Strategies of qualitative inquiry* (2nd Ed.) (pp. 1 - 45). Thousand Oaks, CA: Sage.

Denzin, N. K., & Lincoln, Y. S. (2005). Introduction: The discipline and practice of qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (3rd ed., pp. 1–32). Sage.

Denzin, N. K., & Lincoln, Y. S. (2018). *SAGE handbook of qualitative research*. Los Angeles: SAGE Publications Ltd.

Denzin, N. K., Lincoln, Y. S., Giardina, M. D., & Cannella, G. S. (Eds.). (2024). *The SAGE handbook of qualitative research* (6th ed.). Sage.

Derrida, J. (1967). *Of grammatology*. Baltimore, MD: Johns Hopkins University Press.

Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational researcher*, 32(1), 5-8.

Dixon-Woods, M., Cavers, D., Agarwal, S., Annandale, E., Arthur, A., Harvey, J., Hsu, R., Katbamna, S., Olsen, R., Smith, L., Riley, R., & Sutton, A. J. (2006). Conducting a critical interpretive synthesis of the literature on access to healthcare by vulnerable groups. *BMC Medical Research Methodology*, 6, 35. <https://doi.org/10.1186/1471-2288-6-35>

Elmholdt, K. T., Gill, M., & Nielsen, J. A. (in press). "How Many Interviews Do I Need?" An Examination of Interview Numbers and Sampling Moves in Qualitative Research. *Organizational Research Methods*.

El-Nasr, M. S., Durga, S., Shiyko, M., & Sceppa, C. (2015). Data-driven retrospective interviewing (DDRI): a proposed methodology for formative evaluation of pervasive games. *Entertainment computing*, 11, 1-19.

Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2024). GPTs are GPTs: Labor market impact potential of LLMs. *Science*, 384, 1306–1308.

Emerson, R. M., Fretz, R. I., & Shaw, L. L. (2011). *Writing ethnographic fieldnotes* (2nd ed.). University of Chicago Press.

Evans-Pritchard, E. E. (1940). *The Nuer: A description of the modes of livelihood and political institutions of a Nilotic people*. Clarendon Press.

Fanous, A., Goldberg, J., Agarwal, A., Lin, J., Zhou, A., Xu, S., ... & Koyejo, S. (2025, October). Syceval: Evaluating llm sycophancy. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (Vol. 8, No. 1, pp. 893-900).

Farmer, P. (2004). An anthropology of structural violence. *Current Anthropology*, 45(3), 305–325. <https://doi.org/10.1086/382250>

Feigenbaum, E. A. (1977). The art of artificial intelligence: Themes and case studies of knowledge engineering. *Proceedings of the International Joint Conference on Artificial Intelligence*, 1014–1029.

Fielding, N. G. (2008). The role of computer-assisted qualitative data analysis: Impact on emergent methods in qualitative research. In *The Handbook of Emergent Methods* (No. 32, pp. 675-696). Guilford Press.

Fine, M. (1994). Working the hyphens: Reinventing self and other in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 70–82). Sage.

Finlay, L. (2011). *Phenomenology for therapists: Researching the lived world*. Wiley-Blackwell.

Freire, P. (1970). *Pedagogy of the Oppressed*. Continuum Books.

Freund, Y. (2013) Artificial intelligence vs intelligence amplification. California Institute for Telecommunications and Information Technology.
<https://www.youtube.com/watch?v=sEGdszE86bY>

Friard, O., & Gamba, M. (2016). BORIS: a free, versatile open-source event-logging software for video/audio coding and live observations. *Methods in Ecology and Evolution*.

Friedl, J. (2006). *Mastering regular expressions*. O'Reilly Media, Inc.

Friese, S. (2014). *Qualitative Data Analysis with ATLAS. Ti*. SAGE.

Friese, S. (2026). From coding to conversation: A new methodological framework for AI-assisted qualitative analysis. *Qualitative Inquiry*. <https://doi.org/10.1177/10778004251412871>

Friese, S., Nguyen-Trung, K., Powell, S., & Morgan, D. (2025). Beyond Binary Positions: Making Space for Critical and Reflexive GenAI Integration in Qualitative Research. Manuscript under review.

Fu, J., Zhao, G., Deng, Y., Mi, Y., & Qian, X. (2024). Learning to paraphrase for alignment with LLM preference. In *Findings of the Association for Computational Linguistics: EMNLP 2024* (pp. 2394–2407). Association for Computational Linguistics.
<https://aclanthology.org/2024.findings-emnlp.134/>

- Gallifant, J., Afshar, M., Ameen, S., et al. (2025). The TRIPOD-LLM reporting guideline for studies using large language models. *Nature Medicine*, 31, 60–69.
- Garrison, D. R., Cleveland-Innes, M., Koole, M., & Kappelman, J. (2006). Revisiting methodological issues in transcript analysis: Negotiated coding and reliability. *The internet and higher education*, 9(1), 1-8.
- Gee, J. P. (1993). *An introduction to human language: Fundamental concepts in linguistics*.
- Gee, J. P. (1999). *An introduction to discourse analysis: Theory and method*. Routledge.
- Gee, J. P. (2000). Chapter 3: Identity as an analytic lens for research in education. *Review of research in education*, 25(1), 99-125.
- Geertz, C. (1973). *The interpretation of cultures: Selected essays*. Basic Books.
- Geiecke, F., & Jaravel, X. (2024). Conversations at scale: Robust ai-led interviews with a simple open-source platform. Available at SSRN 4974382.
- George, A. L., & Bennett, A. (2005). *Case Studies and Theory Development in the Social Sciences*. MIT Press.
- Giorgi, A. (2009). *The descriptive phenomenological method in psychology: A modified Husserlian approach*. Duquesne University Press.
- Glaser, B., & Strauss, A. (1967/1999). *Discovery of Grounded Theory: Strategies for Qualitative Research* (1st ed.). Routledge. <https://doi.org/10.4324/9780203793206>.
- Goffman, E. (1956). *The Presentation of Self in Everyday Life*. Doubleday
- González, N., Moll, L. C., & Amanti, C. (Eds.). (2005). *Funds of Knowledge: Theorizing Practices in Households, Communities, and Classrooms*. Lawrence Erlbaum Associates.
- Good, B. J. (1994). *Medicine, rationality and experience: An anthropological perspective*. Cambridge University Press.
- Goujon, V., & Ricci, D. (2024). “Shoggoth with Smiley Face”: Knowing-how and letting-know by analogy in artificial intelligence research. *Hybrid. Revue des arts et médiations humaines*, (12).
- Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political analysis*, 21(3), 267-297.
- Gu, J., Jiang, X., Shi, Z., Tan, H., Zhai, X., Xu, C., Li, W., Shen, Y., Ma, S., Liu, H., Wang, S., Zhang, K., Lin, Z., Zhang, B., Ni, L., Gao, W., Wang, Y., & Guo, J. (2024). A survey on LLM-as-a-judge. *The Innovation*, 101253.

Guo, D., Yang, D., Zhang, H., Song, J., Wang, P., Zhu, Q., ... & He, Y. (2025). Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *Nature*, 645, 633–638.

Haensch, A. C., Weiß, B., Steins, P., Chyrva, P., & Bitz, K. (2022). The semi-automatic classification of an open-ended question on panel survey motivation and its application in attrition analysis. *Frontiers in Big Data*, 5, 880554.

Hammersley, M., & Atkinson, P. (2007). *Ethnography: Principles in practice* (3rd ed.). Routledge.

Hayes, A. S. (2025). "Conversing" with qualitative data: Enhancing qualitative research through large language models (LLMs). *International Journal of Qualitative Methods*, 24, 1–19. <https://doi.org/10.1177/16094069251322346>

Heidegger, M. (1927) *Being and Time*. London, UK: SCM Press.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world?. *Behavioral and brain sciences*, 33(2-3), 61-83.

Herr, K., & Anderson, G. L. (2005). *The Action Research Dissertation: A Guide for Students and Faculty*. SAGE.

Heston, T. F., & Gillette, J. (2025). Large Language Models Demonstrate Distinct Personality Profiles. *Cureus*, 17(5).

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist*, 41(2), 111–127. https://doi.org/10.1207/s15326985ep4102_4

Hine, C. (2015). *Ethnography for the Internet: Embedded, Embodied and Everyday* (1st ed.). Routledge. <https://doi.org/10.4324/9781003085348>

Hollands, F. & Bakir, I. (2015), "Efficiency of Automated Detectors of Learner Engagement and Affect Compared with Traditional Observation Methods," Center for Benefit-Cost Studies of Education, Teachers College, Columbia University.

Horkheimer, M. (1972). *Critical theory: Selected essays*. Continuum.

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative health research*, 15(9), 1277–1288.

Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., ... & Liu, T. (2025). A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *ACM Transactions on Information Systems*, 43(2), 1-55.

Huang, S., Carter, S., Eaton, J., Pollack, S., Callender, D., III, Makagiansar, N., Gonzalez, M., Carr, S., Hong, J., Handa, K., McCain, M., Millar, T., Julapalli, M., Yun, G., Alt, A., Larsson, C., Leibrock, J., Gallivan, M., Summers, T., Durmus, E., Kearney, M., Shen, J. H., Clark, J., Stern, M.,

& Ganguli, D. (2026, March 18). What 81,000 people want from AI. Anthropic.
<https://anthropic.com/features/81k-interviews>

Husserl, E. (1970). *The crisis of European sciences and transcendental phenomenology: An introduction to phenomenological philosophy* (D. Carr, Trans.). Northwestern University Press.

Hutchins, E. (1995). *Cognition in the Wild*. Cambridge, MA USA: MIT press.

Hutchinson, H., Mackay, W., Westerlund, B., Bederson, B. B., Druin, A., Plaisant, C., Beaudouin-Lafon, M., Conversy, S., Evans, H., Hansen, H., Roussel, N., & Eiderbäck, B. (2003). Technology probes: Inspiring design for and with families. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 17–24). ACM.
<https://doi.org/10.1145/642611.642616>

Hutt, S., Baker, R.S., Ocumpaugh, J., Munshi, A., Andres, J.M.A.L., Karumbaiah, S., Slater, S., Biswas, G., Paquette, L., Bosch, N., van Velsen, M. (2023) Quick Red Fox: An App Supporting a New Paradigm in Qualitative Research on AIED for STEM. In Ouyang, F., Jiao, P., McLaren, B.M., Alavi, A.H. (Eds.) *Artificial Intelligence in STEM Education: The Paradigmatic Shifts in Research, Education, and Technology*.

Janus (2022) Simulators.

https://www.google.com/url?q=https://www.lesswrong.com/posts/vJFdjgzmcXMhNTsx&sa=D&source=docs&ust=1771817617517855&usg=AOvVaw1Cf-maJmW4bHzqCy_J-Skh

Jowsey, T., Braun, V., Clarke, V., Lupton, D., & Fine, M. (2025). We reject the use of generative artificial intelligence for reflexive qualitative research. *Qualitative Inquiry*, 10778004251401851.

Kapania, S., Agnew, W., Eslami, M., Heidari, H., & Fox, S. E. (2025, April). Simulacrum of stories: Examining large language models as qualitative research participants. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (pp. 1-17).

King, G., Keohane, R. O., & Verba, S. (1994). *Designing social inquiry: Scientific inference in qualitative research*. Princeton University Press.

Kleinman, A. (1980). *Patients and healers in the context of culture: An exploration of the borderland between anthropology, medicine, and psychiatry*. University of California Press.

Kleinman, A. (1988). *The illness narratives: Suffering, healing, and the human condition*. Basic Books.

Koskinen, I., Zimmerman, J., Binder, T., Redström, J., & Wensveen, S. (2013). *Design research through practice: From the lab, field, and showroom*. Morgan Kaufmann.

Kozinets, R. V. (2019). *Netnography: The Essential Guide to Qualitative Social Media Research*. Sage. <https://doi.org/10.4324/9781003001430-2>

Krakowski, S., Luger, J., & Raisch, S. (2023). Artificial intelligence and the changing sources of competitive advantage. *Strategic Management Journal*, 44(6), 1425-1452.

Kreis, R. (2017). #refugeesnotwelcome: Anti-refugee discourse on Twitter. *Discourse & Communication*, 11(5), 498–514.

Krippendorff, K. (2018). *Content Analysis: An Introduction to Its Methodology* (4th ed.). Sage.

Krippendorff, K. (2019). *Content analysis*. SAGE Publications, Inc.,
<https://doi.org/10.4135/9781071878781>

Krumm, A. E., Means, B., & Bienkowski, M. (2018). *Learning Analytics Goes to School: A Collaborative Approach to Improving Education*. New York, NY: Routledge

Kupiec, T. (2024). *AI-Supported Individual Interviews: Opportunities and Threats for Evaluation*. Unpublished manuscript, accessed 12 April 2026
from https://www.researchgate.net/profile/Tomasz-Kupiec-3/publication/391366554_AI-Supported_Individual_Interviews_Opportunities_and_Threats_for_Evaluation/links/6813c3d8df0e3f544f504e35/AI-Supported-Individual-Interviews-Opportunities-and-Threats-for-Evaluation.pdf

Lakatos, I. (1970). Falsification and the Methodology of Scientific Research Programmes. In *Criticism and the Growth of Knowledge: Volume 4: Proceedings of the International Colloquium in the Philosophy of Science*, London. Cambridge University Press.

Lakatos, I. (1978). *The methodology of scientific research programmes: Philosophical papers* (Vol. 1). Cambridge University Press.

Lakoff, G., & Johnson, M. (1980). The metaphorical structure of the human conceptual system. *Cognitive science*, 4(2), 195-208.

Laughing Phd Mba, Sophy & Ruvalcaba, Bo. (2025). No, AI Doesn't Drink a Bottle of Water per prompt. *The Engineering Reality of AI Infrastructure. Closed-Loop AI Cooling and Power* | Technical Report. 10.13140/RG.2.2.31382.25921.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge university press.

Lawrence, C. (2026) Meet Rachel: the AI agent that phoned 3,000 pubs to price a pint. *Tech EU*. <https://tech.eu/2026/03/20/meet-rachel-the-ai-agent-that-phoned-3000-pubs-to-price-a-pint/>

Lewin, K. (1946). Action research and minority problems. *Journal of Social Issues*, 2(4), 34–46.
<https://doi.org/10.1111/j.1540-4560.1946.tb02295.x>

Li, D., Li, L., & Qiu, H. S. (2025). ChatGPT is not A Man but Das Man: Representativeness and Structural Consistency of Silicon Samples Generated by Large Language Models. *arXiv preprint arXiv:2507.02919*.

- Li, K., Liu, T., Bashkansky, N., Bau, D., Viégas, F., Pfister, H., & Wattenberg, M. (2024). Measuring and controlling instruction (in) stability in language model dialogs. arXiv preprint arXiv:2402.10962.
- Lin, Z. (2025). Six fallacies in substituting large language models for human participants. *Advances in Methods and Practices in Psychological Science*, 8(3), 25152459251357566.
- Lincoln, Y.S. and Guba, E.G. (1985) *Naturalistic Inquiry*. SAGE, Thousand Oaks
- Liu, N. F., Lin, K., Hewitt, J., Paranjape, A., Bevilacqua, M., Petroni, F., & Liang, P. (2024a). Lost in the middle: How language models use long contexts. *Transactions of the association for computational linguistics*, 12, 157-173.
- Liu, Z., Dai, J., Conati, C., & McGrenere, J. (2025a). Envisioning AI support during semi-structured interviews across the expertise spectrum. *Proceedings of the ACM on Human-Computer Interaction*, 9(2), 1-29.
- Lopez-Fierro, S., & Nguyen, H. (2024). Making Human-AI contributions transparent in qualitative coding. In *Proceedings of the 17th International Conference on Computer-Supported Collaborative Learning-CSCCL 2024*, pp. 3-10. International Society of the Learning Sciences.
- Lu, Y., Yao, B., Gu, H., Huang, J., Wang, Z. J., Li, Y., ... & Wang, D. (2025, April). Uxagent: An llm agent-based usability testing framework for web design. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (pp. 1-12).
- Lupton, D. (2015). *Digital sociology*. Routledge.
- Malinowski, B. (1922/1978). *Argonauts of the Western Pacific: An account of native enterprise and adventure in the Archipelagoes of Melanesian New Guinea*. Routledge.
- Mani Adhikari, D., Hartland, A., Weber, I., & Cannanure, V. K. (2025, July). Exploring llms for automated generation and adaptation of questionnaires. In *Proceedings of the 7th ACM Conference on Conversational User Interfaces* (pp. 1-23).
- Marmot, M. (2005). Social determinants of health inequalities. *The Lancet*, 365(9464), 1099–1104. [https://doi.org/10.1016/S0140-6736\(05\)71146-6](https://doi.org/10.1016/S0140-6736(05)71146-6)
- Martin, P., & Bateson, P. (2007). *Measuring Behaviour: An Introductory Guide* (3rd ed.). Cambridge University Press.
- Maxwell, J. A. (2013). *Qualitative research design: An interactive approach*. Sage.
- Mayring, Philipp (2000). Qualitative Content Analysis [28 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 1(2), Art. 20, <http://nbnresolving.de/urn:nbn:de:0114-fqs0002204>.

Mead, G. H. (1934). *Mind, self & society from the standpoint of a social behaviorist*. Chicago, Ill: The University of Chicago Press.

Mead, M. (1928). *Coming of age in Samoa: A psychological study of primitive youth for Western civilisation*. William Morrow.

Meherab, M. M., Billah, M. M., Rahman, K. S., Sharmin, L., Islam, T., Mahmud, Z. Z., ... & Bhuiyan, T. (2026). Advancing NLP Equity: A Secondary Benchmark Evaluation of Multilingual Language Models for Underrepresented Languages. In *Second Workshop on Language Models for Underserved Communities (LM4UC)*.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis*. Sage.

Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill.

Mountz, A., Bonds, A., Mansfield, B., Loyd, J., Hyndman, J., Walton-Roberts, M., et al. (2015). For slow scholarship: A feminist politics of resistance through collective action in the neoliberal university. *ACME: An International Journal for Critical Geographies*, 14(4), 1235–1259.

Moustakas, C. (1994). *Phenomenological research methods*. Sage.

Mowshowitz, Z. (2025). OpenAI model differentiation 101. Don't Worry About the Vase (Substack). <https://thezvi.substack.com/p/openai-model-differentiation-101>

Mwatkins, Rumbelow, J. (2023) SolidGoldMagikarp II: technical details and more recent findings. Available online at <https://www.lesswrong.com/posts/Ya9LzwEbfaAMY8ABo/solidgoldmagikarp-ii-technical-details-and-more-recent>

Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review*, 83(4), 435–450

Newell, A., & Simon, H. A. (1961). Computer Simulation of Human Thinking: A theory of problem solving expressed as a computer program permits simulation of thinking processes. *Science*, 134(3495), 2011-2017.

Nielsen, J. (1994). *Usability engineering*. Morgan Kaufmann.

Nöth, W. (1990). *Handbook of semiotics*. Indiana University Press.

O'Connor, C., & Joffe, H. (2020). Intercoder reliability in qualitative research: Debates and practical guidelines. *International Journal of Qualitative Methods*, 19.

Ocuppaugh, J., Baker, R.S., Rodrigo, M.M.T., Salvi, A. van Velsen, M., Aghababayan, A., Martin, T. (2015). HART: The Human Affect Recording Tool. *Proceedings of the ACM Special Interest Group on the Design of Communication (SIGDOC)*.

Ohsaki, A., & Kaneko, D. (2025, October). Computer-Assisted Code Generation Using Combination of Generative Artificial Intelligence, Stepwise Coding, and Topic Modeling. In International Conference on Quantitative Ethnography (pp. 302-317). Cham: Springer Nature Switzerland.

Olsson, C., Elhage, N., Nanda, N., Joseph, N., DasSarma, N., Henighan, T., ... & Olah, C. (2022). In-context learning and induction heads. arXiv preprint arXiv:2209.11895.

Packer, M. J. (2011). The science of qualitative research. Cambridge University Press.

Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health and Mental Health Services Research*, 42(5), 533–544.

Park, J. S., O'Brien, J. C., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023). Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23)* (pp. 1–22). Association for Computing Machinery. <https://doi.org/10.1145/3586183.3606763>

Parker, J. L., Richard, V., & Becker, K. (2023a). Flexibility & Iteration: Exploring the Potential of Large Language Models in Developing and Refining Interview Protocols. *The Qualitative Report*, 28(9), 2772-2790.

Parker, J. L., Richard, V. M., & Becker, K. (2023). Guidelines for the Integration of Large Language Models in Developing and Refining Interview Protocols. *The Qualitative Report*, 28(12), 3460–3474.

Patton, M. Q. (2002). *Qualitative research & evaluation methods* (3rd ed.). SAGE.

Peng, S., Kalliamvakou, E., Cihon, P., & Demirer, M. (2023). The impact of ai on developer productivity: Evidence from github copilot. arXiv preprint arXiv:2302.06590.

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The Development and Psychometric Properties of LIWC2015*. University of Texas at Austin.

Perrigo, B. (2023). The new AI-powered Bing is threatening users. That's no laughing matter. *TIME*. <https://time.com/6256529/bing-openai-chatgpt-danger-alignment/>

Pierson, P. (2004). *Politics in Time: History, Institutions, and Social Analysis*. Princeton University Press.

Pilny, A., McAninch, K., Slone, A., & Moore, K. (2019). Using supervised machine learning in automated content analysis: An example using relational uncertainty. *Communication Methods and Measures*, 13(4), 287–304.

Polkinghorne, D. E. (1995). Narrative configuration in qualitative analysis. *International Journal of Qualitative Studies in Education*, 8(1), 5–23.

Pope, C., & Mays, N. (Eds.). (2006). *Qualitative research in health care* (3rd ed.). Blackwell Publishing.

Potekhin, A. (2024). Harnessing ChatGPT for business model validation via AI-simulated interviews. Unpublished manuscript, Karelia University of Applied Sciences.

Pringle, E. (2023). Microsoft's ChatGPT-powered Bing is becoming a pushy pick-up artist that wants you to leave your partner: 'You're married, but you're not happy'. *Fortune*.
<https://fortune.com/2023/02/17/microsoft-chatgpt-bing-romantic-love/>

Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2023, July). Robust speech recognition via large-scale weak supervision. In *International conference on machine learning* (pp. 28492–28518). PMLR.

Ratner et al. (2017) – Snorkel: Rapid Training Data Creation with Weak Supervision

Ravitch, S. M., & Carl, N. M. (2019). *Qualitative research: Bridging the conceptual, theoretical, and methodological*. Sage publications.

Reimers, N., & Gurevych, I. (2019, November). Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)* (pp. 3982-3992).

Rettberg, J. W., & Wigers, H. (2025). AI-generated stories favour stability over change: homogeneity and cultural stereotyping in narratives generated by gpt-4o-mini. *arXiv preprint arXiv:2507.22445*.

Riessman, C. K. (2008). *Narrative methods for the human sciences*. Sage.

Robertson, A. (2024, February 21). Google apologizes for “missing the mark” after Gemini generated racially diverse Nazis. *The Verge*.
<https://www.theverge.com/2024/2/21/24079371/google-ai-gemini-generative-inaccurate-historical>

Roddenberry, G. (1964) *Star Trek*. Los Angeles, CA: NBC.

Russell, S., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4), 696-735.

Saldaña, J. (2011). *Fundamentals of qualitative research*. Oxford University Press.

Sartre, J-P. (1946) Existentialism is Humanism. New Haven, CT USA: Yale University Press.

Scheper-Hughes, N., & Lock, M. M. (1987). The mindful body: A prolegomenon to future work in medical anthropology. *Medical Anthropology Quarterly*, 1(1), 6–41.
<https://doi.org/10.1525/maq.1987.1.1.02a00020>

Schreier, M. (2012). *Qualitative content analysis in practice*. (Vols. 1-0). SAGE Publications Ltd, <https://doi.org/10.4135/9781529682571>

Schroeder, H., Aubin Le Quéré, M., Randazzo, C., Mimno, D., & Schoenebeck, S. (2025, April). Large language models in qualitative research: uses, tensions, and intentions. In *Proceedings of the 2025 CHI conference on human factors in computing systems* (pp. 1-17).

Seawright, J., & Gerring, J. (2008). Case selection techniques in case study research: A menu of qualitative and quantitative options. *Political research quarterly*, 61(2), 294-308.

Seidel, J. V. (1998). Qualitative data analysis. In *The Ethnograph v5.0: A user's guide* (Appendix E). Qualis Research.

Settles, B. (1995). Active Learning Literature Survey. *Science*, 10(3), 237-304.

Shaffer, D. W. (2017). *Quantitative ethnography*. Cathcart Press.

Shaffer, D. W., & Ruis, A. R. (2021, January). How we code. In *International Conference on Quantitative Ethnography* (pp. 62-77). Cham: Springer International Publishing.

Shaffer, D. W., & Ruis, A. R. (2025, December). Models all the way down. In *Advances in Quantitative Ethnography: ICQE 2025 Proceedings*. Springer.

Shaffer, D. W., Collier, W., & Ruis, A. R. (2016). A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of Learning Analytics*, 3(3), 9–45.

Shortliffe, E. H. (1976). *Computer-Based Medical Consultations: MYCIN*. Elsevier.

Silver, C., & Lewins, A. (2014). *Using Software in Qualitative Research: A Step-by-Step Guide*. SAGE.

Smith, J. A. (2004). Reflecting on the development of interpretative phenomenological analysis and its contribution to qualitative research in psychology. *Qualitative research in psychology*, 1(1), 39-54.

Smith, J. A. (Ed.). (2004). *Qualitative psychology: A practical guide to research methods*. Sage.

Smith, J. A., Flowers, P., & Larkin, M. (2009). *Interpretative phenomenological analysis: Theory, method and research*. Sage.

Spradley, J. P. (1979). *The ethnographic interview*. Holt, Rinehart and Winston.

St. Amant, R. (2014) *The Use of Tools*. North Carolina State University Department of Computer Science Technical Report TR-2014-4.

Stake, R. E. (1995). *The art of case study research*. Sage.

Stephenson, N. (1994). *Snow crash*. Penguin UK.

Strubell, E., Ganesh, A., & McCallum, A. (2019, July). Energy and policy considerations for deep learning in NLP. In *Proceedings of the 57th annual meeting of the association for computational linguistics* (pp. 3645-3650).

Swidler, A. (1986). Culture in action: Symbols and strategies. *American Sociological Review*, 51(2), 273–286. <https://doi.org/10.2307/2095521>

Tang, K. S. (2025). AI-textuality: Expanding intertextuality to theorize human-AI interaction with generative artificial intelligence. *Applied Linguistics*, article amaf016.

Tao, Y., Viberg, O., Baker, R. S., & Kizilcec, R. F. (2024). Cultural bias and cultural alignment of large language models. *PNAS Nexus*, 3(9), pgae346.

Tavory, I., & Timmermans, S. (2014). *Abductive analysis: Theorizing qualitative research*. University of Chicago press.

Tracy, S. J. (2010). Qualitative quality: Eight “big-tent” criteria for excellent qualitative research. *Qualitative Inquiry*, 16(10), 837–851.

Tuhiwai Smith, L. (1999). *Decolonizing Methodologies: Research and Indigenous Peoples*. Zed Books.

Turpin, M., Michael, J., Perez, E., & Bowman, S. (2023). Language models don't always say what they think: Unfaithful explanations in chain-of-thought prompting. *Advances in Neural Information Processing Systems*, 36, 74952-74965.

Van Dis, E. A., Bollen, J., Zuidema, W., Van Rooij, R., & Bockting, C. L. (2023). ChatGPT: Five priorities for research. *Nature*, 614(7947), 224–226.

van Manen, M. (1990). *Researching lived experience: Human science for an action sensitive pedagogy*. State University of New York Press.

Vindrola-Padros, C., & Johnson, G. A. (2020). Rapid techniques in qualitative research: a critical review of the literature. *Qualitative health research*, 30(10), 1596-1604.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wactlar, H. D., Kanade, T., Smith, M. A., & Stevens, S. M. (2002). Intelligent access to digital video: Informedia project. *Computer*, 29(5), 46-52.

Wang, A., Morgenstern, J., & Dickerson, J. P. (2025). Large language models that replace human participants can harmfully misportray and flatten identity groups. *Nature Machine Intelligence*, 7(3), 400-411.

Wedel, M., & Kamakura, W. A. (2000). *Market segmentation: Conceptual and methodological foundations*. Springer Science & Business Media.

Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., ... & Fedus, W. (2022a) Emergent Abilities of Large Language Models. *Transactions on Machine Learning Research*.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., ... & Zhou, D. (2022b). Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35, 24824-24837.

Wei, Z., Baker, R.S., Pankiewicz, M., Chen, Y. (2026) Using an AI Interviewer to Investigate Student Cognition in Introductory Programming Error Resolution. *EdArxiv Preprint 46nhx_v2*.

Wen, C., Clough, P., Paton, R., & Middleton, R. (2025). Leveraging large language models for thematic analysis: A case study in the charity sector. *AI & Society*, 1–18.

Wickelgren, W. A. (1995). *How to solve mathematical problems*. Courier Corporation.

Willig, C. (2013). *Introducing Qualitative Research in Psychology* (3rd ed.). Open University Press.

Willis, J. W., Jost, M., & Nilakanta, R. (2007). *Foundations of qualitative research: Interpretive and critical approaches*. Sage.

Wise, A. F., Zhao, Y., & Hausknecht, S. N. (2014). Learning analytics for online discussions: Embedded and extracted approaches. *Journal of Learning Analytics*, 1(2), 48–71.

Worsley, M. (2018). Multimodal learning analytics for the qualitative researcher. *Proceedings of the International Conference of the Learning Sciences (ICLS)*.

Wutich, A., Beresford, M., & Bernard, H. R. (2024). Sample Sizes for 10 Types of Qualitative Data Analysis: An Integrative Review, Empirical Guidance, and Next Steps. *International Journal of Qualitative Methods*, 23, 16094069241296206.

Wuttke, A., Aßenmacher, M., Klamm, C., Lang, M., & Kreuter, F. (2025, May). 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)* (pp. 179-204).

Xiao, Z., Zhou, M. X., Liao, Q. V., Mark, G., Chi, C., Chen, W., & Yang, H. (2020). Tell me about yourself: Using an AI-powered chatbot to conduct conversational surveys with open-ended questions. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 27(3), 1-37.

Xu, Z., Liu, Y., Deng, G., Li, Y., & Picek, S. (2024, August). A comprehensive study of jailbreak attack versus defense for large language models. In *Findings of the Association for Computational Linguistics: ACL 2024* (pp. 7432-7449).

Yamamoto, A., Koda, M., Ogawa, H., Miyoshi, T., Maeda, Y., Otsuka, F., & Ino, H. (2024). Enhancing medical interview skills through AI-simulated patient interactions: nonrandomized controlled trial. *JMIR medical education*, 10(1), e58753.

Yang, Z., Zhang, Y., Liu, T., Yang, J., Lin, J., Zhou, C., & Sui, Z. (2024). Can large language models always solve easy problems if they can solve harder ones? In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing* (pp. 1531–1555). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2024.emnlp-main.92>

Yardley, L. (2000). Dilemmas in qualitative health research. *Psychology and Health*, 15(2), 215–228.

Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). Sage.

Yuli_Ban (2023) Pink Shoggoths: What does alignment look like in practice?
https://www.lesswrong.com/posts/9y5RpyyFJX4GaqPLC/pink-shoggoths-what-does-alignment-look-like-in-practice?utm_source=chatgpt.com

Zhang, A., Tanzer, G., Marcheret, E., & Mortensen, D. R. (2024). Shortcomings of LLMs for Low-Resource Translation: Retrieval and Understanding are Both the Problem. *arXiv preprint arXiv:2406.15625*.

Zhang, H., Liu, Y., Guan, X., Cai, J., & Carroll, J. M. (2025). Harnessing the Power of AI in Qualitative Research: Role Assignment, Engagement, and User Perceptions of AI-Generated Follow-Up Questions in Semi-Structured Interviews. *arXiv preprint arXiv:2509.12709*.

Zheng, L., Chiang, W. L., Sheng, Y., Zhuang, S., Wu, Z., Zhuang, Y., ... & Stoica, I. (2023). Judging llm-as-a-judge with mt-bench and chatbot arena. *Advances in neural information processing systems*, 36, 46595-46623.

Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. *Journal of Educational Psychology*, 81(3), 329–339. <https://doi.org/10.1037/0022-0663.81.3.329>

Zook, M., Barocas, S., Boyd, D., Crawford, K., Keller, E., Gangadharan, S. P., ... & Pasquale, F. (2017). Ten simple rules for responsible big data research. *PLoS computational biology*, 13(3), e1005399.

Zörgő, S., Swiecki, Z., & Ruis, A. R. (2021). Exploring the effects of segmentation on semi-structured interview data with epistemic network analysis. In International conference on quantitative ethnography (pp. 78–90). Springer.