**References**
**Version 1.08 October 26, 2026**


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

AI Weekly. (2025). Palmer Lucky explains his ChatGPT hack to get it to do ANYTHING [Video]. YouTube. 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), Article e0300024.

Andriushchenko, M., Croce, F., & Flammarion, N. (2023). Jailbreaking leading safety-aligned LLMs with simple adaptive attacks. 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. In Proceedings of the International Conference on Computers in Education.

Astudillo, G. (2025). A large language model approach for in-depth qualitative text analysis. In Proceedings of the 15th IEEE International Conference on Pattern Recognition Systems (ICPRS). IEEE. https://doi.org/10.1109/ICPRS66293.2025.11302863

Atil, B., Aykent, S., Chittams, A., Fu, L., Passonneau, R. J., Radcliffe, E., Baldwin, B., et al. (2024). Non-determinism of "deterministic" LLM settings. arXiv. https://arxiv.org/abs/2408.04667

Bai, Y., Jones, A., Ndousse, K., Askell, A., Chen, A., DasSarma, N., Kaplan, J., et al. (2022). Training a helpful and harmless assistant with reinforcement learning from human feedback. arXiv. https://arxiv.org/abs/2204.05862

Baker, R. S. (2024). Big data and education (8th ed.). University of Pennsylvania.

Baker, R. S., & Hawn, M. A. (2022). Algorithmic bias in education. International Journal of Artificial Intelligence in Education, 32, 1052–1092.

Baker, R.S., Hutt, S. (2025) MORF: A Post-Mortem. *Proceedings of the International Conference on Learning Analytics and Knowledge*.

Baker, R. S., Hutt, S., Bosch, N., Ocumpaugh, J., Biswas, G., Paquette, L., Munshi, A., et al. (2024). Detector-driven classroom interviewing: Focusing qualitative researcher time by selecting cases in situ. Educational Technology Research and Development, 72(5), 2841–2863.

Baker, R.S., Mills, C., Lan, A., Lehman, B., Barany, A. (2026) A Taylorizable Process for Textual Detector Development for Research-Practice Partnerships. *Journal of Learning Engineering, 1* (1).

Baker, R. S., Ocumpaugh, J., & Andres, J. M. A. L. (2020). BROMP quantitative field observations: A review. In R. Feldman (Ed.), Learning science: Theory, research, and practice.

Baker, R., Siemens, G. (2022) Educational data mining and learning analytics. In Sawyer, K. (Ed.) *Cambridge Handbook of the Learning Sciences: 3rd Edition*.

Bakharia, A., Shibani, A., Lim, L. A., McCluskey, T., & Shum, S. B. (2025). From transcripts to themes: A trustworthy workflow for qualitative analysis using large language models. In CEUR Proceedings of the Workshop 'From Data to Discovery: LLMs for Qualitative Analysis in Education', 15th International Learning Analytics & Knowledge Conference (LAK '25).

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., et al. (2024, July). ChatGPT for education research: Exploring the potential of large language models for qualitative codebook development. In Proceedings of the International Conference on Artificial Intelligence in Education (pp. 134–149). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-64299-9_10

Barany, A., Shah, M., & Foster, A. (2021, January). Connecting curricular design and student identity change: An epistemic network analysis. In Proceedings of the International Conference on Quantitative Ethnography (pp. 155–169). Springer International Publishing.

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

Baumann, J., Röttger, P., Urman, A., Wendsjö, A., Plaza-del-Arco, F. M., Gruber, J. B., & Hovy, D. (2025). Large language model hacking: Quantifying the hidden risks of using llms for text annotation. *arXiv preprint arXiv:2509.08825*.

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

Beck, C. T. (2004). Method for studying same phenomenon from different lenses: A methodological review. Qualitative Health Research, 14(10), 1431–1442.

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

Belur, J., Tompson, L., Thornton, A., & Simon, M. (2021). Interrater reliability in systematic review methodology: Exploring variation in coder decision-making. Sociological Methods & Research, 50(2), 837–865.

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.

Beresford, M., Wutich, A., du Bray, M. V., Ruth, A., Stotts, R., SturtzSreetharan, C., & Brewis, A. (2022). Coding qualitative data at scale: Guidance for large coder teams based on 18 studies. International Journal of Qualitative Methods, 21. https://doi.org/10.1177/16094069221075860

Berger, P. L., & Luckmann, T. (1966). The social construction of reality: A treatise in the sociology of knowledge. Doubleday.

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.

Biesta, G., & Burbules, N. C. (2003). Pragmatism and educational research. Rowman & Littlefield.

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, Article e59050.

Birks, M., Chapman, Y., & Francis, K. (2008). Memoing in qualitative research: Probing data and processes. Journal of Research in Nursing, 13(1), 68–75.

Birt, L., Scott, S., Cavers, D., Campbell, C., & Walter, F. (2016). Member checking: A tool to enhance trustworthiness or merely a nod to validation? Qualitative Health Research, 26(13), 1802–1811.

Blair, A. (2025). Gemini 3 is evaluation-paranoid and contaminated. LessWrong. https://www.lesswrong.com/posts/8uKQyjrAgCcWpfmcs/gemini-3-is-evaluation-paranoid-and-contaminated

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, 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., et al. (2021). On the opportunities and risks of foundation models. arXiv. https://arxiv.org/abs/2108.07258

Booth, R. (2026, February 4). From "nerdy" Gemini to "edgy" Grok: How developers are shaping AI behaviours. The Guardian. https://www.theguardian.com/technology/2026/feb/03/gemini-grok-chatgpt-claude-qwen-ai-chatbots-identity-crisis

Borchers, C., Shahrokhian, B., Balzan, F., Tajik, E., Sankaranarayanan, S., & Simon, S. (2025a). Temperature and persona shape LLM agent consensus with minimal accuracy gains in qualitative coding. arXiv. https://arxiv.org/abs/2507.11198

Borchers, C., Zhang, J., Fleischer, H., Schanze, S., Aleven, V., & Baker, R. (2025b). Large language models generalize SRL prediction to new languages within but not between domains. *Journal of Educational Data Mining*, *17*(2), 24-54.

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 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.

Braidotti, R. (2013). The posthuman. John Wiley & Sons.

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. https://doi.org/10.1191/1478088706qp063oa

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. https://doi.org/10.1080/2159676X.2019.1628806

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 be(com)ing 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.

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). Sage.

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. OER Commons. https://oercommons.org/courseware/lesson/114039/overview

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

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

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

Cai, Z., Eagan, B., Marquart, C., & Shaffer, D. W. (2023). LSTM neural network assisted regex development for qualitative coding. In C. Damşa & A. Barany (Eds.), Advances in quantitative ethnography: ICQE 2022 proceedings (pp. 17–29). Springer.

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 Advances in quantitative ethnography: ICQE 2023 proceedings (pp. 455–469). Springer.

Campbell, J. L., Quincy, C., Osserman, J., & Pedersen, O. K. (2013). Coding in-depth semistructured interviews: Problems of unitization and intercoder reliability and agreement. Sociological Methods & Research, 42(3), 294–320.

Cao, J., Hao, Z., & Yu, J. (2026, June). Decoding Student Dialogue: A Multi-dimensional Comparison and Bias Analysis of Large Language Models as Annotation Tools. In *International Conference on Artificial Intelligence in Education* (pp. 378-386). Cham: Springer Nature Switzerland.

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

Chinh, B., Zade, H., Ganji, A., & Aragon, C. (2019, May). Ways of qualitative coding: A case study of four strategies for resolving disagreements. In Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (pp. 1–6).

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 Advances in quantitative ethnography: ICQE 2022 proceedings. Springer.

Chopra, F., & Haaland, I. (2024). Conducting qualitative interviews with AI (CESifo Working Paper No. 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. In Concise encyclopedia of philosophy of language and linguistics (pp. 85–87).

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

Clarke, A. E., Washburn, R., & Friese, C. (Eds.). (2015). Situational analysis in practice: Mapping research with grounded theory (1st ed.). Routledge.

Clarke, S. N., Sushil, S., Dennis, K., Lee, U. S., Gomoll, A., & Gates, Z. (2023). Developing shared ways of seeing data: The perils and possibilities of achieving intercoder agreement. International Journal of Qualitative Methods, 22.

Cleo Nardo. (2023). The Waluigi effect (mega-post). LessWrong. 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.

Cole, R. (2024). Inter-rater reliability methods in qualitative case study research. *Sociological Methods & Research, 53*(4), 1944–1975. https://doi.org/10.1177/00491241231156971

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

Comte, A. (1853). The positive philosophy of Auguste Comte (H. Martineau, Trans.). Chapman. (Original work published 1830–1842)

Conrad, P. (1985). The meaning of medications: Another look at compliance. Social Science & Medicine, 20(1), 29–37.

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). Sage.

Crenshaw, K. W., Gotanda, N., Peller, G., & Thomas, K. (1995). Critical race theory: The key writings that formed the movement [Faculty Books, 101].

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

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

Crotty, M. (1998). The foundations of social research: Meaning and perspective in the research process. Sage.

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., Scurrell, J. V., Brown, E. M., Entenmann, J., & Daepp, 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.

Cui, W., Hao, J., Jiang, Y., Kyllonen, P. C., & Kerzabi, E. (2026, March). Automated coding of communication data using large language models: a comparison of hierarchical and direct prompting strategies. In *Frontiers in Education* (Vol. 11, p. 1764154). Frontiers Media SA.

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), Article e2517912122.

D'Ignazio, C., & Klein, L. F. (2020). *Data Feminism.* MIT Press. https://doi.org/10.7551/mitpress/11805.001.0001

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. https://doi.org/10.1007/s11135-024-01950-6

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 Working Paper No. 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. Informatics, 10(4), Article 78.

Denzin, N. K. (1978). The research act: A theoretical introduction to sociological methods (2nd ed.). McGraw-Hill.

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). 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. (Eds.). (2005). The Sage handbook of qualitative research (3rd ed.). Sage Publications Ltd.

Denzin, N. K., & Lincoln, Y. S. (2018). SAGE handbook of qualitative research. Sage.

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. Johns Hopkins University Press.

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

Dewey, J. (1938). Experience and education. Macmillan.

Dickson-Swift, V., James, E. L., Kippen, S., & Liamputtong, P. (2009). Researching sensitive topics: Qualitative research as emotion work. Qualitative Research, 9(1), 61–79.

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, Article 35.

Dunivin, Z. O. (2025). Scaling hermeneutics: A guide to qualitative coding with LLMs for reflexive content analysis. EPJ Data Science, 14, Article 28. https://doi.org/10.1140/epjds/s13688-025-00548-8

Dunivin, Z. O., Noori, M., Frey, S., & Atkinson, C. (2026). Self-reflection in Automated Qualitative Coding: Improving Text Annotation through Secondary LLM Critique. arXiv preprint arXiv:2601.09905.

Durkheim, E. (1982). The rules of sociological method (W. D. Halls, Trans.). Free Press. (Original work published 1895)

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.

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.

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.

Feenberg, A. (2002). Transforming technology: A critical theory revisited. Oxford University Press.

Fehring, L., Frings, J., Rust, P., Kempny, C., Thürmann, P. A., & Meister, S. (2025). Extension of the Consolidated Criteria for Reporting Qualitative Research Guideline to Large Language Models (COREQ+LLM): Protocol for a multiphase study. *JMIR Research Protocols, 14*, e78682. https://doi.org/10.2196/78682[GC1]

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

Fine, M., & Torre, M. E. (2019). Critical participatory action research: A feminist project for validity and solidarity. Psychology of Women Quarterly, 43(4), 433–444. https://doi.org/10.1177/0361684319865255

Finlay, L. (2002). "Outing" the researcher: The provenance, process, and practice of reflexivity. Qualitative Health Research, 12(4), 531–545.

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

Flick, U. (2018). Triangulation in data collection. In The Sage handbook of qualitative data collection.

Foucault, M. (1980). Power/knowledge: Selected interviews and other writings 1972–1977. Pantheon Books.

Freire, P. (1970). Pedagogy of the oppressed. Continuum.

Freund, Y. (2013). Artificial intelligence vs intelligence amplification [Video]. YouTube. 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.

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).

Gadamer, H.-G. (2004). Truth and method (2nd rev. ed., J. Weinsheimer & D. G. Marshall, Trans.). Continuum. (Original work published 1960)

Gale, N. K., Heath, G., Cameron, E., Rashid, S., & Redwood, S. (2013). Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC medical research methodology*, *13*(1), 117.

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.

Garces, K., Fernandez-Nieto, G. M., Zhao, L., Samaraweera, S., Gasevic, D., Martinez-Maldonado, R., & Echeverria, V. (2026, June). Scalable LLM-based Coding of Dialogue in Healthcare Simulation: Balancing Coding Performance, Processing Time, and

Environmental Impact. In Proceedings of the Thirteenth ACM Conference on Learning@ Scale (pp. 79-90).

Garcia, N. M., Onyewuenyi, A. C., & Shaffer, D. W. (2025, October). Critical Quantitative Ethnography (CritQE): A Pathway Toward Convergence Research and Equity. In *International Conference on Quantitative Ethnography* (pp. 52-66). Cham: Springer Nature Switzerland.

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. SSRN. https://ssrn.com/abstract=4974382

George, A. L., & Bennett, A. (2005). Case studies and theory development in the social sciences. MIT Press.

Gillespie, A. (2026). Provoking interpretation: Using large language models as devil's advocates. *Qualitative Psychology*.

Giorgi, A. (2009). The descriptive phenomenological method in psychology: A modified Husserlian approach. Duquesne University Press.

Glaser, B. G., & Strauss, A. L. (1967). The discovery of grounded theory: Strategies for qualitative research. Aldine.

Glaser, B., & Strauss, A. (1967/1999). Discovery of grounded theory: Strategies for qualitative research (1st ed.). Routledge.

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).

Greene, J. A., & Azevedo, R. (2009). A macro-level analysis of SRL processes and their relations to the acquisition of a sophisticated mental model of a complex system. Contemporary Educational Psychology, 34(1), 18–29.

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, Article 101253.

Guba, E. G., & Lincoln, Y. S. (1989). Fourth generation evaluation. Sage.

Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), Handbook of qualitative research (pp. 105–117). Sage.

Guillemin, M., & Gillam, L. (2004). Ethics, reflexivity, and "ethically important moments" in research. Qualitative Inquiry, 10(2), 261–280. https://doi.org/10.1177/1077800403262360

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, Article 880554.

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

Hao, J., Cui, W., Kyllonen, P., Kerzabi, E., Liu, L., & Flor, M. (2024). Scaling up the evaluation of collaborative problem solving: Promises and challenges of coding chat data with ChatGPT. arXiv preprint arXiv:2411.10246.

Haraway, D. (1988). Situated knowledges: The science question in feminism and the privilege of partial perspective. Feminist Studies, 14(3), 575–599.

Harding, S. G. (Ed.). (1987). Feminism and methodology: Social science issues. Indiana University Press.

Hayes, A. S. (2025). "Conversing" with qualitative data: Enhancing qualitative research through large language models (LLMs). International Journal of Qualitative Methods, 24, 1–19.

Heidegger, M. (1927). Being and time. SCM Press.

Heidegger, M. (1962). Being and time (J. Macquarrie & E. Robinson, Trans.). Blackwell. (Original work published 1927)

Heidegger, M. (1999). Ontology—The hermeneutics of facticity (J. van Buren, Trans.). Indiana University Press. (Original work published 1923)

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.

Hill, C. E., Thompson, B. J., & Williams, E. N. (1997). A guide to conducting consensual qualitative research. The Counseling Psychologist, 25(4), 517–572.

Hill, J.L., Bellocchi, A., Male, S. & Cebon, P. (2026) Analysing emotions in engineering students' reflections using appraisal theory through an artificial intelligence tool, European Journal of Engineering Education, 51:5, 1336-1358, DOI: 10.1080/03043797.2025.2579777

Hine, C. (2015). Ethnography for the internet: Embedded, embodied and everyday (1st ed.). Routledge.

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.

Holmes, A. G. D. (2020). Researcher positionality — A consideration of its influence and place in qualitative research: A new researcher guide. Shanlax International Journal of Education, 8(4), 1–10.

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

Hou, C., Zhu, G., Zheng, J., Zhang, L., Huang, X., Zhong, T., & Ker, C. L. (2024, March). Prompt-based and fine-tuned GPT models for context-dependent and -independent deductive coding in social annotation. In Proceedings of the 14th Learning Analytics and Knowledge Conference (pp. 518–528).

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., et al. (2026, March 18). What 81,000 people want from AI. Anthropic. https://anthropic.com/features/81k-interviews

Hulley, S. B., Cummings, S. R., Browner, W. S., Grady, D. G., & Newman, T. B. (2007). Designing clinical research (3rd ed.). Lippincott Williams & Wilkins.

Husserl, E. (1982). Ideas pertaining to a pure phenomenology and to a phenomenological philosophy—First book: General introduction to a pure phenomenology (F. Kersten, Trans.). Martinus Nijhoff. (Original work published 1913)

Hutchins, E. (1995). Cognition in the wild. 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.

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 F. Ouyang, P. Jiao, B. M. McLaren, & A. H. Alavi (Eds.), Artificial intelligence in STEM education: The paradigmatic shifts in research, education, and technology.

Janus. (2022). Simulators. LessWrong. https://www.lesswrong.com/posts/vJFdjigzmcXMhNTsx

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. Educational Researcher, 33(7), 14–26. https://doi.org/10.3102/0013189X033007014

Jones, T. (2025)

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. https://doi.org/10.1177/10778004251401851

Kambhampati, S., Stechly, K., Valmeekam, K., Saldyt, L., Bhambri, S., Palod, V., Gundawar, A., et al. (2025). Stop anthropomorphizing intermediate tokens as reasoning/thinking traces! arXiv. https://arxiv.org/abs/2504.09762

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).

Katz, A., Fleming, G. C., & Main, J. B. (2026). Thematic analysis with open-source generative AI and machine learning: A new method for inductive qualitative codebook development. *Humanities and Social Sciences Communications*, *13*(1), 209.

Kemmis, S. (2007). Critical theory and participatory action research. In The Sage handbook of action research (pp. 121–138). Sage.

Kim, J., & Mitchell, J. (2026). How AI coders discuss, disagree, and reach consensus: Challenges and opportunities for LLM-based qualitative coding. arXiv. https://arxiv.org/abs/2609.11109

Kincheloe, J. L., & McLaren, P. (2005). Rethinking critical theory and qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), The Sage handbook of qualitative research (3rd ed., pp. 303–342). Sage.

Kindon, S., Pain, R., & Kesby, M. (2007). Participatory action research: Origins, approaches and methods. In Participatory action research approaches and methods (pp. 35–44). Routledge.

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.

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.

Krumm, A. E., Means, B., & Bienkowski, M. (2018). Learning analytics goes to school: A collaborative approach to improving education. Routledge.

Kuhn, T. S. (1970). The structure of scientific revolutions (2nd ed.). University of Chicago Press. (Original work published 1962)

Kupiec, T. (2024). AI-supported individual interviews: Opportunities and threats for evaluation [Unpublished manuscript]. ResearchGate.

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), Criticism and the growth of knowledge (pp. 91–196). Cambridge University Press.

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

Lather, P. (1986). Issues of validity in openly ideological research: Between a rock and a soft place. Interchange, 17(4), 63–84. https://doi.org/10.1007/BF01807017

Lather, P. (1993). Fertile obsession: Validity after poststructuralism. The Sociological Quarterly, 34(4), 673–693.

Laudan, L. (1977). Progress and its problems: Towards a theory of scientific growth. University of California Press.

Laughing PhD MBA, S., & Ruvalcaba, B. (2025). No, AI doesn't drink a bottle of water per prompt. The engineering reality of AI infrastructure [Technical report]. https://doi.org/10.13140/RG.2.2.31382.25921

Lave, J., & Wenger, E. (1991). Situated learning: Legitimate peripheral participation. Cambridge University Press.

Lawrence, C. (2026, March 20). 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/

Lee, S., Zhou, J., Ao, C., Li, K., Du, X., He, S., ... & Ni, S. (2025, July). Quantification of large language model distillation. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 4985-5004).

Lewin, K. (1946). Action research and minority problems. Journal of Social Issues, 2(4), 34–46.

Li, H., Kuang, H., & Botelho, A. F. (2026a). Can We Trust AI's Self-assessment? Evaluating and Improving LLM Confidence Calibration in Educational Dialogue Coding. In *International Conference on Artificial Intelligence in Education* (pp. 262-276). Cham: Springer Nature Switzerland.

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. https://arxiv.org/abs/2507.02919

Li, F., Mason, M. L., Levin, D. T., & Wise, A. F. (2026b). When can we trust AI coding of student-generated text? a committee-based approach to diagnosing agreement and uncertainty at scale. In *International Conference on Artificial Intelligence in Education* (pp. 370-379). Cham: Springer Nature Switzerland.

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. https://arxiv.org/abs/2402.10962

Li, Z., Tang, L., Xia, M., Li, X., Chen, N., Gašević, D., & Fan, Y. (2026c). When LLMs Fall Short in Deductive Coding: Model Comparisons and Human–AI Collaboration Workflow Design. In *Proceedings of the LAK26: 16th International Learning Analytics and Knowledge Conference* (pp. 685-696).

Lin, Z. (2025). Six fallacies in substituting large language models for human participants. Advances in Methods and Practices in Psychological Science, 8(3).

Lincoln, Y. S., & Denzin, N. K. (Eds.). (2003). Turning points in qualitative research: Tying knots in a handkerchief (Vol. 2). Bloomsbury Publishing.

Lincoln, Y. S., & Guba, E. G. (1985). Naturalistic inquiry. Sage.

Liu, J., Shen, D., Zhang, Y., Dolan, B., Carin, L., & Chen, W. (2022). What makes good in-context examples for GPT-3? In Proceedings of Deep Learning Inside Out (DeeLIO 2022) (pp. 100–114). Association for Computational Linguistics.

Liu, N. F., Lin, K., Hewitt, J., Paranjape, A., Bevilacqua, M., Petroni, F., & Liang, P. (2024). 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.

Liu, X., Zambrano, A. F., Baker, R. S., Barany, A., Ocumpaugh, J., Zhang, J., Pankiewicz, M., Nasiar, N., & Wei, Z. (2025b). Qualitative coding with GPT-4: Where it works better. Journal of Learning Analytics, 12(1), 169–185.

Liu, X., Zhou, Y., Ocumpaugh, J., Barany, A., Zambrano, A.F., Wei, Z., Baker, R.S., Giordano, C. (2025c) Not All Who Wander Are Lost: Trailblazing Trajectories in a Minecraft-Based Learning Environment. *Proceedings of the International Conference on Quantitative Ethnography*.

Liu, Z., Mondol, A., Jiao, X., Chao, J., Li, L., & Xing, W. (2026, June). From Examples to Rules? Exploring Inductive Reverse Engineering and Deductive Few-Shot Coding via LLMs for Qualitative Data Analysis. In *International Conference on Artificial Intelligence in Education* (pp. 639-653). Cham: Springer Nature Switzerland.

Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2002). Content analysis in mass communication: Assessment and reporting of intercoder reliability. Human Communication Research, 28(4), 587–604.

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 – CSCL 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 Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (pp. 1–12).

Lupo, L., Magnusson, O., Hovy, D., Naurin, E., & Wängnerud, L. (2023). Towards human-level text coding with LLMs: The case of fatherhood roles in public policy documents. *arXiv preprint arXiv:2311.11844*.

Lupton, D. (2015). Digital sociology. Routledge.

MacPhail, C., Khoza, N., Abler, L., & Ranganathan, M. (2016). Process guidelines for establishing intercoder reliability in qualitative studies. Qualitative Research, 16(2), 198–212.

MacQueen, K. M., McLellan, E., Kay, K., & Milstein, B. (1998). Codebook development for team-based qualitative analysis. Cultural Anthropology Methods, 10(2), 31–36.

Madsen, A., Chandar, S., & Reddy, S. (2024). Are self-explanations from large language models faithful? In Findings of the Association for Computational Linguistics: ACL 2024. Association for Computational Linguistics.

Malinowski, B. (1922/1978). Argonauts of the Western Pacific: An account of native enterprise and adventure in the Archipelagoes of Melanesian New Guinea. Routledge.

Malterud, K., Siersma, V. D., & Guassora, A. D. (2016). Sample size in qualitative interview studies: Guided by information power. Qualitative Health Research, 26(13), 1753–1760.

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.

Marrone, R., Barthakur, A., Randall, D, Mottaghi, N., Kovanovic, V., de Laat, M. (in press) Evaluating Generative AI as a Supportive Analytic Partner in Qualitative Coding of Metacognitive Student Reflections. To appear in *Journal of Educational Data Mining.*

Martin, P., & Bateson, P. (2007). Measuring behaviour: An introductory guide (3rd ed.). Cambridge University Press.

Maxwell, J. A. (1992). Understanding and validity in qualitative research. Harvard Educational Review, 62(3), 279–301. https://doi.org/10.17763/haer.62.3.8323320856251826

Maxwell, J. A. (2013). Qualitative research design: An interactive approach. Sage.

Mayring, P. (2000). Qualitative content analysis. Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, 1(2), Article 20.

McClure, J., Smyslova, D., Hall, A., & Jiang, S. (2024). Deductive coding's role in AI vs. human performance. In Proceedings of the 17th International Conference on Educational Data Mining (pp. 809-813).

McDonald, N., Schoenebeck, S., & Forte, A. (2019). Reliability and inter-rater reliability in qualitative research: Norms and guidelines for CSCW and HCI practice. Proceedings of the ACM on Human-Computer Interaction, 3(CSCW), Article 72.

Mead, G. H. (1934). Mind, self and society from the standpoint of a social behaviorist. 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: A methods sourcebook (3rd ed.). Sage.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2020). Qualitative data analysis: A methods sourcebook (4th ed.). Sage.

Milner, H. R., IV. (2007). Race, culture, and researcher positionality: Working through dangers seen, unseen, and unforeseen. Educational Researcher, 36(7), 388–400. https://doi.org/10.3102/0013189X07309471

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

Montgomery, P., & Bailey, P. H. (2007). Field notes and theoretical memos in grounded theory. *Western Journal of nursing research*, *29*(1), 65-79.

Morgan, D. L. (2007). Paradigms lost and pragmatism regained: Methodological implications of combining qualitative and quantitative methods. Journal of Mixed Methods Research, 1(1), 48–76. https://doi.org/10.1177/2345678906292462

Morse, J. M. (1998). Validity by committee. Qualitative Health Research, 8(4), 443–445.

Morse, J. M., Barrett, M., Mayan, M., Olson, K., & Spiers, J. (2002). Verification strategies for establishing reliability and validity in qualitative research. International Journal of Qualitative Methods, 1(2), 13–22. https://doi.org/10.1177/160940690200100202

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. LessWrong. 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.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. Psychological Review, 84(3), 231–259.

Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., & Vazire, S. (2022). Replicability, robustness, and reproducibility in psychological science. Annual Review of Psychology, 73, 719–748. https://doi.org/10.1146/annurev-psych-020821-114157

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.

Ocumpaugh, J., Baker, R. S., Rodrigo, M. M. T., Salvi, A., van Velsen, M., Aghababyan, A., & Martin, T. (2015). HART: The Human Affect Recording Tool. In 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 Proceedings of the International Conference on Quantitative Ethnography (pp. 302–317). 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. https://arxiv.org/abs/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.

Parfenova, A., Marfurt, A., Pfeffer, J., & Denzler, A. (2025, April). Text annotation via inductive coding: Comparing human experts to LLMs in qualitative data analysis. In Findings of the Association for Computational Linguistics: NAACL 2025 (pp. 6456–6469).

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 (pp. 1–22).

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.

Parker, M. J., Anderson, C., Stone, C., & Oh, Y. (2025). A large language model approach to educational survey feedback analysis. International journal of artificial intelligence in education, 35(2), 444-481.

Patton, M. Q. (1999). Enhancing the quality and credibility of qualitative analysis. Health Services Research, 34(5, Pt. 2), 1189–1208.

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

Patton, M. Q. (2015). Qualitative research & evaluation methods: Integrating theory and practice (4th ed.). Sage.

Peng, S., Kalliamvakou, E., Cihon, P., & Demirer, M. (2023). The impact of AI on developer productivity: Evidence from GitHub Copilot. arXiv. https://arxiv.org/abs/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, February 16). The new AI-powered Bing is threatening users. That's no laughing matter. TIME. https://time.com/6256529/bing-openai-chatgpt-danger-alignment/

Phillips, D. C., & Burbules, N. C. (2000). Postpositivism and educational research. Rowman & Littlefield.

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.

Popper, K. (1959). The logic of scientific discovery. Basic Books.

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

Pringle, E. (2023, February 17). 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/

Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others. Personality and Social Psychology Bulletin, 28(3), 369–381.

Qiao, T., Walker, C., Cunningham, C., & Koh, Y. S. (2025, April). Thematic-LM: A LLM-based multi-agent system for large-scale thematic analysis. In Proceedings of the ACM on Web Conference 2025 (pp. 649–658).

Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2023, July). Robust speech recognition via large-scale weak supervision. In Proceedings of the International Conference on Machine Learning (pp. 28492–28518).

Ramanathan, S., Lim, L. A., Mottaghi, N. R., & Buckingham Shum, S. (2025, March). When the prompt becomes the codebook: Grounded Prompt Engineering (GROPROE) and its application to belonging analytics. In *Proceedings of the 15th international learning analytics and knowledge conference* (pp. 713-725).

Ratajczyk, E., Brady, U., Baggio, J. A., Barnett, A. J., Perez-Ibarra, I., Rollins, N., Rubiños, C., et al. (2016). Challenges and opportunities in coding the commons: Problems, procedures, and potential solutions in large-N comparative case studies. International Journal of the Commons, 10(2), 440–466.

Ratner, A., De Sa, C., Wu, S., Selsam, D., & Ré, C. (2017). Snorkel: Rapid training data creation with weak supervision. Proceedings of the VLDB Endowment, 11(3), 269–282.

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

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 (pp. 3982–3992).

Ren, S., Nguyen, H., Bernacki, M., Yu, L., & Greene, J. (2026). Using Large Language Models for Automated Coding of Self-Regulated Learning Think-Aloud Protocol Data. *Journal of Learning Analytics*, 1-24.

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. https://arxiv.org/abs/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 [TV series]. NBC.

Rodgers, B. L., & Cowles, K. V. (1993). The qualitative research audit trail: A complex collection of documentation. Research in Nursing & Health, 16(3), 219–226.

Rodríguez Triana, M. J., Saban, M., Asensio-Pérez, J. I., Prieto, L. P., Haba-Ortuo, I., Villa-Torrano, C., & Gillet, D. (2025, September). Code-Aware LLM Prompting in Deductive Qualitative Analysis: A Study in Multi-framework Analysis of Learning Designs. In European Conference on Technology Enhanced Learning (pp. 427-441). Cham: Springer Nature Switzerland.

Russell, S., & Norvig, P. (2021). Artificial intelligence: A modern approach (4th ed.). Pearson.

Ryle, G. (1968). The thinking of thoughts: What is "le penseur" doing? (University Lectures No. 18). University of Saskatchewan.

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.

Saldaña, J. (2021). The coding manual for qualitative researchers (4th ed.). Sage.

Samaraweera, S., Zhao, L., Echeverria, V., Alfredo, R., Chen, G., Davis, J., ... & Dharmaratne, A. (2026, April). From formal learning to professional practice: Automated llm-based coding and visualisation of

team dialogue in in-situ healthcare simulation. In *Proceedings of the LAK26: 16th International Learning Analytics and Knowledge Conference* (pp. 926-932).

Sandelowski, M. (1993). Rigor or rigor mortis: The problem of rigor in qualitative research revisited. Advances in Nursing Science, 16(2), 1–8. https://doi.org/10.1097/00012272-199312000-00002

Sarumi, O. O., Welch, C., & Braun, D. (2026). Fine-grained perspectives: Modeling explanations with annotator-specific rationales. In Proceedings of NLPerspectives@LREC 2026 (pp. 66–75).

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.

Schick, T., Dwivedi-Yu, J., Dessì, R., Lomeli, M., Hambro, E., Zettlemoyer, L., Cancedda, N., & Scialom, T. (2023). Toolformer: Language models can teach themselves to use tools. Advances in Neural Information Processing Systems, 36.

Schleiermacher, F. (1998). Hermeneutics and criticism: And other writings (A. Bowie, Ed. & Trans.). Cambridge University Press.

Schmidt, L. K. (2006). Understanding hermeneutics. Acumen Publishing.

Schooler, J. W., Ohlsson, S., & Brooks, K. (1993). Thoughts beyond words: When language overshadows insight. Journal of Experimental Psychology: General, 122(2), 166–183.

Schreier, M. (2012). Qualitative content analysis in practice. Sage.

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. (2009). Active learning literature survey. Science, 10(3), 237–304.

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

Shaffer, D. W., & Cai, Z. (2024, November). Perfect sampling. In *International Conference on Quantitative Ethnography* (pp. 40-57). Cham: Springer Nature Switzerland.

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.

Shaffer, D. W., & Ruis, A. R. (2021, January). How we code. In Proceedings of the International Conference on Quantitative Ethnography (pp. 62–77). 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.

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.

Sprague, Z., Yin, F., Rodriguez, J. D., Jiang, D., Wadhwa, M., Singhal, P., et al. (2025). To CoT or not to CoT? Chain-of-thought helps mainly on math and symbolic reasoning. In The Thirteenth International Conference on Learning Representations.

St. Amant, R. (2014). The use of tools (Technical Report No. TR-2014-4). North Carolina State University, Department of Computer Science.

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

Steiger, M., Bharucha, T. J., Venkatagiri, S., Riedl, M. J., & Lease, M. (2021). The psychological well-being of content moderators: The emotional labor of commercial moderation and avenues for improving support. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. ACM. https://doi.org/10.1145/3411764.3445092

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.

Tai, R. H., Bentley, L. R., Xia, X., Sitt, J. M., Fankhauser, S. C., Chicas-Mosier, A. M., & Monteith, B. G. (2024). An examination of the use of large language models to aid analysis of textual data. *International Journal of Qualitative Methods*, *23*, 16094069241231168.

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), 346.

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

Theelen, H., Vreuls, J., & Rutten, J. (2024). Doing research with help from ChatGPT: Promising examples for coding and inter-rater reliability. International Journal of Technology in Education (IJTE), 7(1), 1-18. https://doi.org/10.46328/ijte.537

Tracy, S. J. (2010). Qualitative quality: Eight "big-tent" criteria for excellent qualitative research. Qualitative Inquiry, 16(10), 837–851. https://doi.org/10.1177/1077800410383121

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.

Vanacore, K., & Kizilcec, R. (2026, June). How well do large language models recognize instructional moves? Establishing baselines for foundation models in educational discourse. In Proceedings of the Thirteenth ACM Conference on Learning@ Scale (pp. 330-338).

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.

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.

von Glasersfeld, E. (1995). Radical constructivism: A way of knowing and learning. Routledge.

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.

Waelbers, B., Henkel, A. P., & Bromuri, S. (2025, March). Beyond Traditional Quality Monitoring in Call Centers: A Comparative Analysis of Human Evaluators and Large Language Models. In *International Conference on Information Management* (pp. 314-323). Cham: Springer Nature Switzerland.

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, Z., Baker, R. S., Pankiewicz, M., & Chen, Y. (2026). Using an AI interviewer to investigate student cognition in introductory programming error resolution. EdArXiv.

Wei, Z., Liu, X., Barany, A., Ocumpaugh, J., Mehta, S., Nasiar, N., Baker, R.S., Zambrano, A.F., Vanacore, K., Giordano, C. (under review) Cultural Alignment and Biases in Qualitative Coding: Comparing GPT and Human Coders. Manuscript under review.

Wei, Z., Nasiar, N., Zambrano, A.F., Liu,X., Ocumpaugh, J., Barany,A., Baker, R.S., Giordano, C. (2025) Exploring Students' Interest-Driven Patterns of Scientific Observation in Minecraft. *Proceedings of the International Conference of the Learning Sciences*.

Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., & Fedus, W. (2022). Emergent abilities of large language models. Transactions on Machine Learning Research.

Wei, Z., Zhang, J., Baker, R.S., Borchers, C., Barany, A., Liu, X. (2026) Scaffolding Human-LLM Qualitative Coding Through Inference Evaluation and Boundary Definition. *Proceedings of the International Conference on Quantitative Ethnography.*

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. Advances in Neural Information Processing Systems, 35, 24824–24837.

Wellert, L., Braun, A., Becker, S., Frick, P., Brendel, W., & Lachner, A. (2025). Theory-Prompt-Validation: A Practice-Oriented Approach to Using LLMs for Verbal Coding in the Learning Sciences. https://osf.io/download/zyje8/

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.

Wheeldon, J., & Faubert, J. (2009). Framing experience: Concept maps, mind maps, and data collection in qualitative research. *International journal of qualitative methods*, *8*(3), 68-83.

Whitehead, R., Nguyen, A., & Järvelä, S. (2025). Utilizing multimodal large language models for video analysis of posture in studying collaborative learning: A case study. *Journal of Learning Analytics*, *12*(1), 186-200.

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. In Proceedings of the International Conference of the Learning Sciences.

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.

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 (pp. 179–204).

Xiao, Z., Yuan, X., Liao, Q. V., Abdelghani, R., & Oudeyer, P. Y. (2023, March). Supporting qualitative analysis with large language models: Combining codebook with GPT-3 for deductive coding. In Companion proceedings of the 28th international conference on intelligent user interfaces (pp. 75-78).

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, 27(3), 1–37.

Xu, Z., Khatri, V., Dai, Y., Liu, X., Li, S., Zhang, X., & Yu, R. (2026, April). Enhancing LLM-Based Data Annotation with Error Decomposition. In *Proceedings of the LAK26: 16th International Learning Analytics and Knowledge Conference* (pp. 325-335).

Xu, H., Yi, S., Lim, T., Xu, J., Well, A., Mery, C., & Ding, Y. (2025). Tama: A human-AI collaborative thematic analysis framework using multi-agent LLMs for clinical interviews. ACM Transactions on Computing for Healthcare.

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), Article 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).

Yardley, L. (2000). Dilemmas in qualitative health research. Psychology and Health, 15(2), 215–228. https://doi.org/10.1080/08870440008400302

Ye, C., Borman, T. H., & Borman, G. D. (2026). Automating Self-Affirmation Essay Coding: Fine-Tuned BERT Performance Comparable to Human Coders and Comparison with GPT-4. *Journal of Educational Data Mining*, *18*(1), 66-88.

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? LessWrong. https://www.lesswrong.com/posts/9y5RpyyFJX4GaqPLC/pink-shoggoths-what-does-alignment-look-like-in-practice

Zahid, O., Hsiang-Pan, J., Arastoopour Irgens, G., Behboudi, A., & Lane, A. C. (2025, October). Of Humans and Machines: Evaluating the Efficacy of GPT-4 in Coding Discourse Data. In International Conference on Quantitative Ethnography (pp. 117-132). Cham: Springer Nature Switzerland.

Zambrano, A. F., Liu, X., Barany, A., Baker, R. S., Kim, J., & Nasiar, N. (2023). From nCoder to ChatGPT: From automated coding to refining human coding. In Proceedings of the International Conference on Quantitative Ethnography. https://doi.org/10.1007/978-3-031-47014-1_32

Zamecnik, A., de Laat, M., Claassen, A., & Baker, R. (2026) Evaluating Large Language Models in the Assessment of Collaborative Problem-Solving Skills: Evidence from Classroom Dialogue. EdArXiv xh27a_v1.

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. https://arxiv.org/abs/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. https://arxiv.org/abs/2509.12709

Zhao, Z., & Liu, Y. (2025). A Confidence-Diversity Framework for Calibrating AI Judgement in Accessible Qualitative Coding Tasks. *arXiv preprint arXiv:2508.02029*.

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.

Zhou, Y., Kang, J., & Nguyen, H. (2025). Can we trust large language models for video analysis: An exploration of hallucination in multimodal LLMs. In Proceedings of the 19th International Conference of the Learning Sciences (ICLS 2025) (pp. 110–118). International Society of the Learning Sciences.

Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. Journal of Educational Psychology, 81(3), 329–339.

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), Article e1005399.

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