Exploring effects of life-like virtual interviewers on respondents' answers in a smartphone survey

Höhne[1,2], Conrad[3], Neuert[4], & Claassen[1,2]

[1] Leibniz University Hannover

[2] German Center for Higher Education Research and Science Studies (DZHW)

[3] University of Michigan

[4] GESIS – Leibniz Institute for the Social Sciences

General Online Research (GOR) Conference

Cologne (Germany) – February 21 to 23, 2024

Introduction I

- Demand for high-quality data from web surveys (Knowledge Sourcing Intelligence, 2023)
- Cost-efficient and streamlined web surveys replace other survey modes, especially in-person interviews (Schober, 2018)
  - *Even large-scale social surveys, such as ESS, start utilizing web surveys*
- Web surveys may not be good candidates for primary survey mode
  - *Depressed response rates* (Daikeler et al., 2020)
  - *Impede participation for people with low levels of literacy* (Höhne, 2023)
  - *Struggle with achieving high data quality* (Callegaro et al., 2015)
- Absence of interviewers impedes the provision of assistance and the creation of trust, motivation, and engagement

Introduction II

- Advances in communication technology and AI introduce new data collection opportunities
- Fusing elements of interviewer-based and web surveys
  - *Life-like virtual interviewers and self-administration*
- Few studies utilized virtual interviewers (Conrad et al., 2015; Conrad et al., 2020; Lind et al., 2013; Schuetzler et al., 2018)
  - *Somewhat inconclusive results on respondent satisfaction and data quality*
  - *Frequently conducted in lab settings with small samples*
- Web surveys with pre-recorded interviewers have quality benefits (West et al., 2022; Conrad et al., 2023)
  - *More disclosure of sensitive behavior, less rounding, and less error variance*
  - *Respondents connect with pre-recorded interviewers*
  - *Results may also apply to virtual interviewers*

Research Questions (RQs)

- How do virtual interviewers affect …

  … item-nonresponse compared to text-based web surveys? ***(RQ1)***

  … answer quality compared to text-based web surveys? ***(RQ2)***

  … survey evaluations compared to text-based web surveys? ***(RQ3)***

- How are virtual interviewers evaluated …

  … by respondents? ***(RQ4)***

Method: AI of Virtual Interviewers

Image generation
Instruction via descriptive keywords using diffusion models or architected transformers (Zhang et al., 2023)

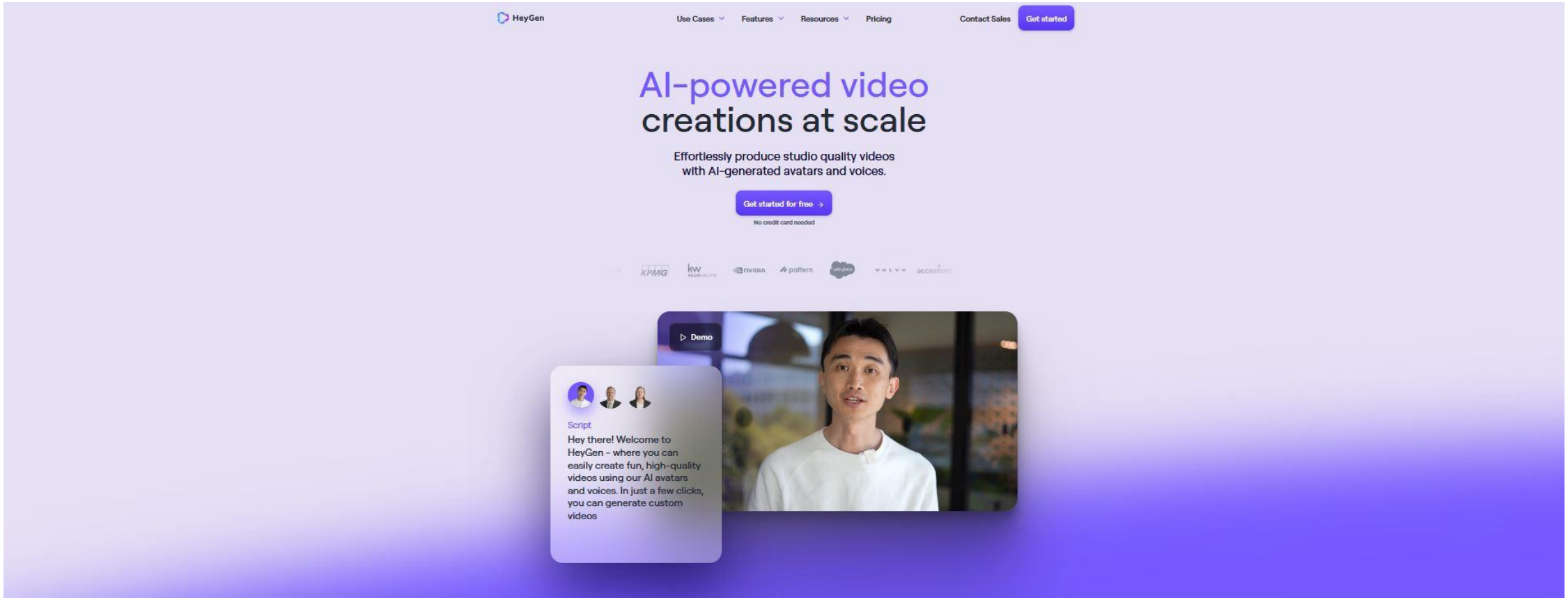
Text generation
Generative Pretrained Transformers and Large Language Models (Vaswani et al., 2017)

Text-to-Speech generation
Statistical Parametric Synthesis and Neural Speech Synthesis (Tan et al., 2021): **1** "text analysis" (heteronyms), **2** "voice parameter prediction" (acoustic model), and **3** "vocoder analysis" (audio snippets)

Speech and image animation
Multistep pipelines of transformers, Recurrent Neural Networks, Convolutional Neural Networks, and Generative Adversarial Networks (Chen et al., 2023)

See Cheung, B. (2023): https://bennycheung.github.io/create-personal-animated-ai-avatar

Method: Creating Virtual Interviewers



See https://www.heygen.com/

Method: Study Design

Male casual
n = (376)

Male business casual
(n = 375)

Female casual
n = (395)

Female business casual
(n = 343)

- Experiment in a smartphone survey (N = 1,871)
- Between-subject design with 5 groups
  - *1 text control without virtual interviewer (n = 382)*
  - *4 treatment groups each with a different virtual interviewer*
- 6 closed questions on women at the workplace (3) and family relations (3)
  - *These questions were adopted from the ESS*
- Respondents had to click ▶ for playing the virtual interviewer video
  - *Respondents were informed that they are surveyed by virtual interviewers*
  - *Videos could be played multiple times*

Method: Example Question Screenshots

Starten wir mit ein paar Fragen zu Job und Karriere.

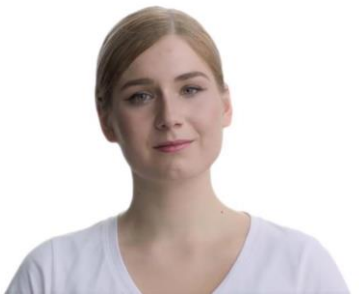
Wie finden Sie es, dass Frauen in Deutschland heutzutage immer häufiger karriereorientiert denken und handeln?

- Sehr gut
- Gut
- Weder gut noch schlecht
- Schlecht
- Sehr schlecht

Text control

- Sehr gut
- Gut
- Weder gut noch schlecht
- Schlecht
- Sehr schlecht

Male casual



- Sehr gut
- Gut
- Weder gut noch schlecht
- Schlecht
- Sehr schlecht

Female casual

Method: Sample

Data collection was conducted in the Respondi/Bilendi panel in Germany in November and December 2023

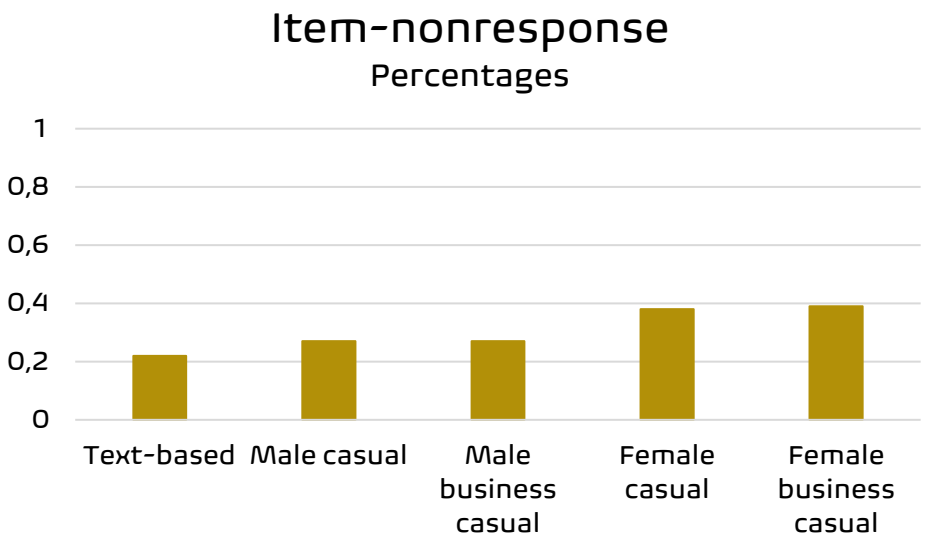
Cross quotas:	Age and gender plus quotas on education
Mean age:	49 years
Gender:	49% females
Education:	44% lower secondary school
	24% intermediate secondary school
	34% at least college preparatory secondary school

Experimental groups do not statistically differ with respect to age, gender, and education

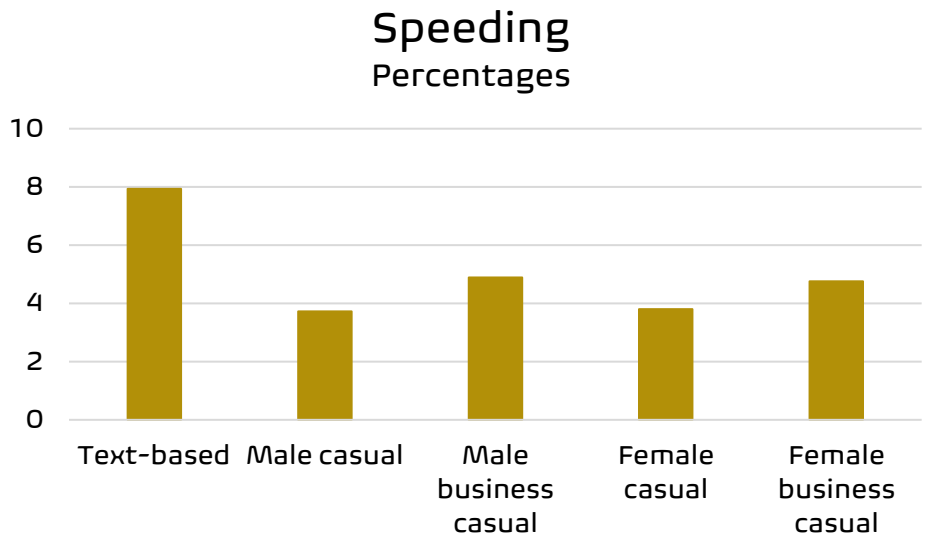
Method: Measures

- RQ1: Determining the share of item-nonresponse – aggregated across all 6 questions
- RQ2: Determining the share of …
  … speeding (the fastest 5% percentile)
  … primacy effects (selecting the 1st option)
  … middle tendency (selecting the 2nd, 3rd, or 4th options)
- RQ3: Determining respondents' survey evaluations in terms of …
  … interest, difficulty, personal feeling, and satisfaction
- RQ4: Determining respondents' virtual interviewer evaluations in terms of …
  … warmth, rapport, naturalness, and authenticity

Results: Research Questions 1 and 2



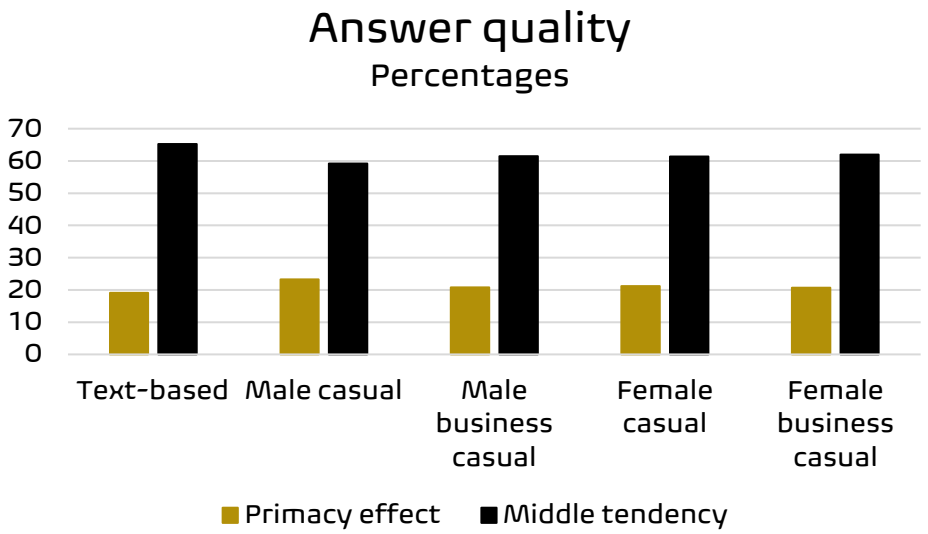
We conducted *z*-tests



We conducted *z*-tests

All conditions with virtual interviewers are less prone to speeding ($p < .05$).

Results: Research Question 2

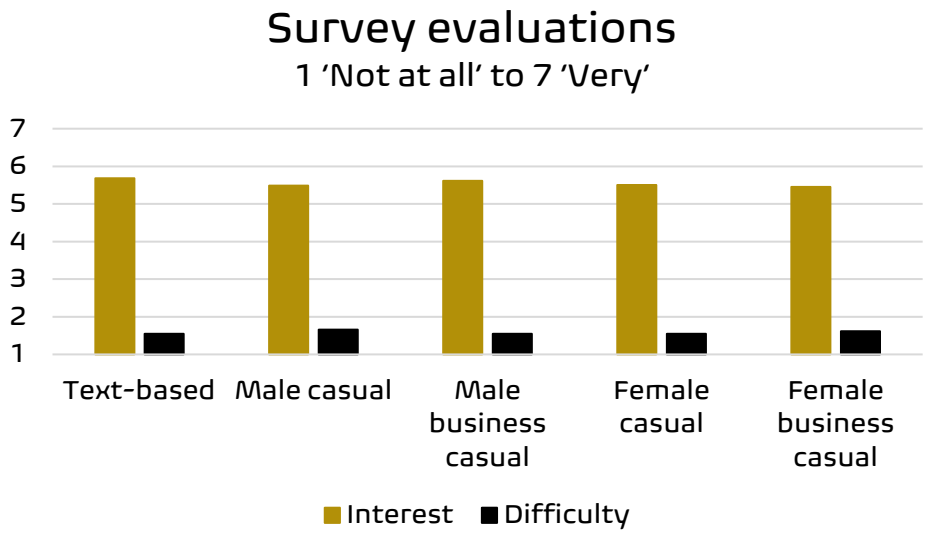


We conducted z-tests

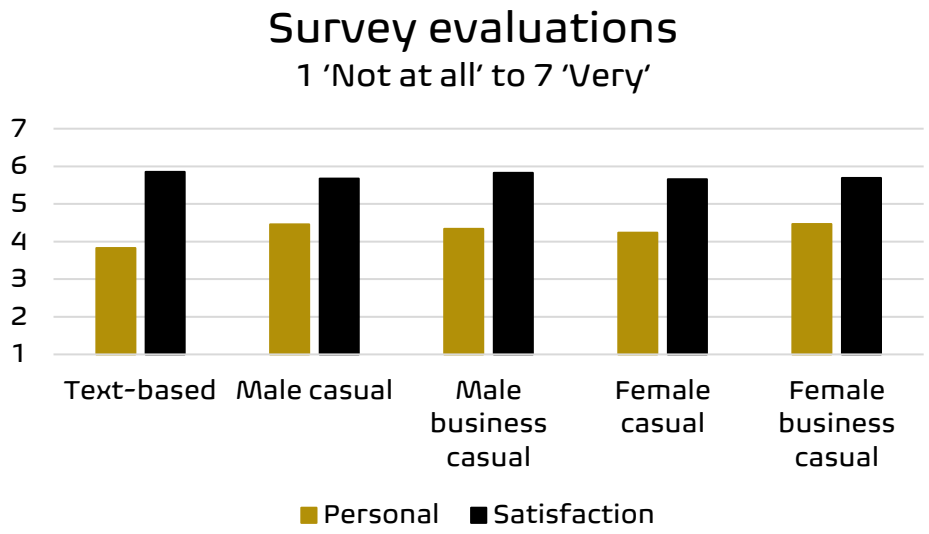
For primacy effects, we find a difference between conditions 1 and 2 ($p < .05$).

All conditions with virtual interviewers are less prone to a middle tendency ($p < .05$).

Results: Research Question 3



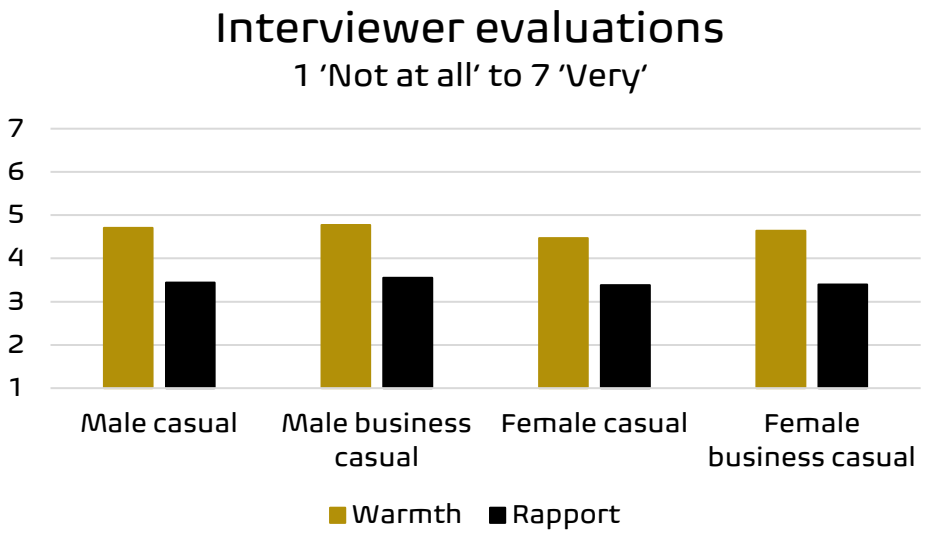
We conducted ANOVAs



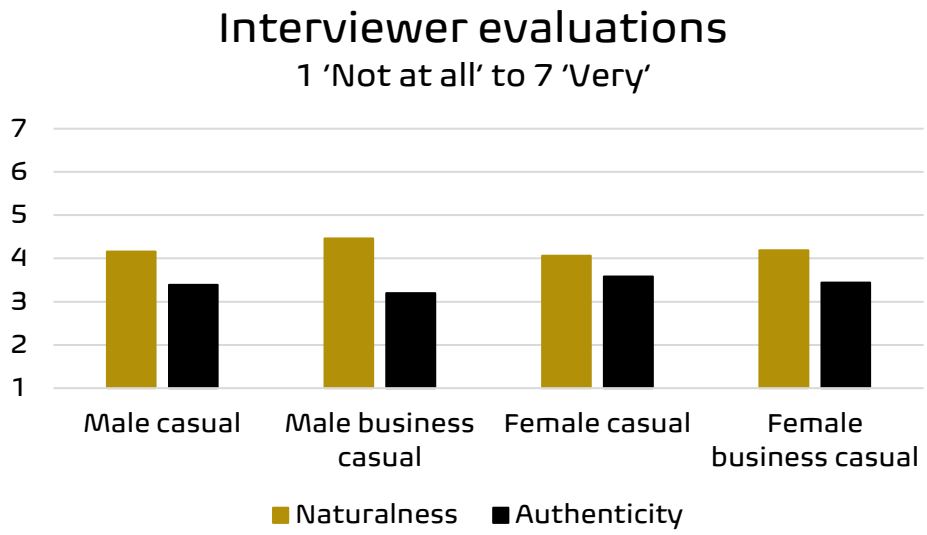
We conducted ANOVAs

All conditions with virtual interviewers are evaluated as being more personal ($p < .05$).

Results: Research Question 4



We conducted ANOVAs



We conducted ANOVAs

Discussion and Conclusion

- Item-nonresponse is low across all conditions (< 1%)
- Text-based condition is more prone to a middle tendency
  - *There are little differences regarding primacy effects*
- Speeding is much more common in the text-based condition
- All virtual interviewer conditions are evaluated as being more personal
  - *Interest and satisfaction ratings are high (> 5 on a 7-point scale)*
- Virtual interviewers perform similarly well in terms of respondent evaluations
  - *Rapport evaluations are somewhat lower than the remaining evaluations*
- Take home message: Virtual interviewers have answer quality benefits and are evaluated well by respondents

Future Avenues

- Investigating data quality and subgroups
  - *Relating speeding to answer quality*
  - *Socially desirable answer behavior and further quality indicators*
- Investigating interviewer assignment strategies
  - *Interviewer selection by respondents and demographic matching*
- Including more natural survey communications
  - *For example, asking respondents for voice (or spoken) answers*
  - *Conversational skills based on paradata and answers*
- Moving from avatars to responsive and autonomous agents
- Growing beyond commercial providers
  - *Tailoring virtual interviewers to survey purposes*

Many thanks for your attention!

www.jkhoehne.eu
@jkhoehne

Literature I

- Callegaro, M., Lozar Manfreda, K., & Vehovar, V. (2015). Web survey methodology. Sage.
- Chen, L., Wu, Z., Li, R., Bao, W., Ling, J., Tan, X., & Zhao, S. (2023). VAST: Vivify your talking avatar via zero-shot expressive facial style transfer. arXiv. DOI: 10.48550/arXiv.2308.04830
- Conrad, F. G., Schober, M. F., Hupp, A. L., West, B. T., Larsen, K. M., Ong, A. R., & Wang, T. (2023). Video in survey interviews: Effects on data quality and respondent experience. Methods Data Anal., 17(2), 135-170. DOI: 10.12758/mda.2022.13
- Conrad, F. G., Schober, M. F., Nielsen, D., & Reichert, H. (2020). Virtual interviewers, social identities, and survey measurement error. In K. Olson, J. D. Smyth, J. Dykema, A. L. Holbrook, F. Kreuter, & B. T. West (Eds.), Interviewer effects from a total survey error perspective (pp. 149–164). CRC Press.
- Conrad, F. G., Schober, M. F., Jans, M., Orlowski, R. A., Nielsen, D., & Levenstein, R. (2015). Comprehension and engagement in survey interviews with virtual agents. Front. Psychol., 6, 1578. DOI: 10.3389/fpsyg.2015.01578
- Daikeler, J., Bosnjak, M., & Lozar Manfreda, K. (2020). Web versus other survey modes: An updated and extended meta-analysis comparing response rates. J. Surv. Stat. Methodol., 8(3), 513–539. DOI: 10.1093/jssam/smz008
- Höhne, J. K. (2023). Are respondents ready for audio and voice communication channels in online surveys? Int. J. Soc. Res. Methodol., 26(3), 335-342. DOI: 10.1080/13645579.2021.1987121
- Knowledge Sourcing Intelligence. (2023). Global online survey software market size, share, opportunities, COVID 19 impact, and trends by application, by product, and by geography – forecasts from 2023 to 2028. https://www.knowledge-sourcing.com/report/global-online-survey-software-market
- Lind, L. H., Schober, M. F., Conrad, F. G., & Reichert, H. (2013). Why do survey respondents disclose more when computers ask the questions? Public Opin. Q., 77(4), 888–935. DOI: 10.1093/poq/nft038
- Schober, M. F. (2018). The future of face-to-face interviewing. Qual. Assur. Educ., 26(2), 290–302. DOI: 10.1108/QAE-06-2017-0033

Literature II

- Schuetzler, R. M., Giboney, J. S., Grimes, G. M., & Nunamaker, J. F. (2018). The influence of conversational agent embodiment and conversational relevance on socially desirable responding. Decis. Support. Syst., 114, 94-102. DOI: 10.1016/j.dss.2018.08.011
- Tan, X., Qin, T, Soong, F., & Liu, T. Y. (2021). A survey on neural speech synthesis. arXiv. DOI: 10.48550/arXiv.2106.15561
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. In NIPS'17. Curran Associates, Inc. https://dl.acm.org/doi/10.5555/3295222.3295349
- West, B. T., Ong, A. R., Conrad, F. G., Schober, M. F., Larsen, K. M., & Hupp, A. L. (2022). Interviewer effects in live video and prerecorded video interviewing. J. Surv. Stat. Methodol., 10(2), 317-336. DOI: 10.1093/jssam/smab040
- Zhang, C., Zhang, C., Zhang, M., & Kweon, S. (2023). Text-to-image diffusion models in generative AI: A survey. arXiv. DOI: 10.48550/arXiv.2303.07909

Appendix: Questions I

- Women at the workplace
  - *How do you feel about the fact that women in Germany are increasingly thinking and acting in a career-oriented way these days? 1 "Very good," 2 "Good," 3 "Neither good, nor bad," 4 "Bad," and 5 "Very bad"*
  - *To what extent do you agree or disagree with the following statement? Women should be given preferential treatment in applications and promotions if they are equally qualified. 1 "Agree strongly," 2 "Agree somewhat," 3 "Neither agree, nor disagree," 4 "Disagree somewhat," and 5 "Disagree strongly"*
  - *To what extent do you agree or disagree with the following statement? Women exaggerate problems they have at work. 1 "Agree strongly," 2 "Agree somewhat," 3 "Neither agree, nor disagree," 4 "Disagree somewhat," and 5 "Disagree strongly"*

Appendix: Questions II

- Family relations
  - *To what extent do you agree or disagree with the following statement? It is better if the man works full-time, and the woman stays at home and takes care of the household and the children. 1 "Agree strongly," 2 "Agree somewhat," 3 "Neither agree, nor disagree," 4 "Disagree somewhat," and 5 "Disagree strongly"*
  - *To what extent do you agree or disagree with the following statement? Even if both parents are working, it is better if the responsibility for the household and the children lies mainly with the woman. 1 "Agree strongly," 2 "Agree somewhat," 3 "Neither agree, nor disagree," 4 "Disagree somewhat," and 5 "Disagree strongly"*
  - *To what extent do you agree or disagree with the following statement? In a family, the man can also be responsible for the household and the children while the woman works full-time. 1 "Agree strongly," 2 "Agree somewhat," 3 "Neither agree, nor disagree," 4 "Disagree somewhat," and 5 "Disagree strongly"*