

Bachelor Thesis: Systematic Literature Review – Resisting Persuasion of AI

Supervisor: Adrian Hofbauer (adrian.hofbauer@uni-passau.de)

Start date: as soon as possible

Motivation and Goals

AI systems are becoming increasingly persuasive: large language models can match or even exceed humans in changing attitudes and behavior in direct conversation (Salvi et al., 2025), raising the risk that users are steered toward choices they would not endorse on reflection. Across psychology, marketing, IS, and HCI, various measures and strategies have been developed to help people resist persuasion attempts — from forewarning and inoculation to warning messages and serious games (e.g., Xiao & Benbasat, 2015). However, this knowledge is scattered across disciplines and was largely developed for static content (e.g., advertisements, fake news). It remains unclear which measures transfer to real-time, adaptive interaction with conversational AI.

The goal of this thesis is to systematically review the literature on empirically evaluated measures and strategies to resist persuasion, organize them along suitable dimensions (e.g., targeted stage of the persuasion process), assess their transferability to conversational AI, and derive challenges and research gaps for the IS domain.

Required Skills

- Strong interest in human–AI interaction and persuasion
- Good English skills
- Prior experience with systematic literature reviews, or willingness to learn the methodology via the starting literature below

Starting Literature (Topic)

Salvi, F., Horta Ribeiro, M., Gallotti, R., & West, R. (2025). On the conversational persuasiveness of GPT-4. *Nature Human Behaviour*, 9(8), 1645–1653.

Xiao, B., & Benbasat, I. (2015). Designing warning messages for detecting biased online product recommendations: An empirical investigation. *Information Systems Research*, 26(4), 793–811.

Albarracín, D., & Karan, A. (2022). Resistance to Persuasion. In *Oxford Research Encyclopedia of Psychology*. Oxford University Press.

Starting Literature (Method)

Webster, J., & Watson, R. T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly*, 26(2), xiii–xxiii. <https://www.jstor.org/stable/4132319>

Templier, M., & Paré, G. (2015). A framework for guiding and evaluating literature reviews. *Communications of the Association for Information Systems*, 37(1), 6. <https://doi.org/10.17705/1CAIS.03706>