



# Is Auditive Communication with ChatGPT an Effective Means of Building Trust Between People and Machines?

A Quantitative Study

---

Marvin Tessitore | Hakan Arda | Prof. Dr. habil. Nicholas Mueller | Prof. Dr. Karsten Huffstadt

Technical University of Applied Sciences Wuerzburg-Schweinfurt (THWS)

Faculty of Computer Science and Business Informatics

Marvin.tessitore.doktoranden@thws.de

Human-AI Interaction • User Trust • Voice Interface • Technology Acceptance Model • Conversational AI

# About the Presenter

M. Sc. Marvin Tessitore | PhD Candidate, THWS Wuerzburg-Schweinfurt

## Affiliation

- THWS Wuerzburg-Schweinfurt
- Faculty of Computer Science & Business Informatics
- Supervisors: Prof. Dr. habil. Nicholas Mueller and Prof. Dr. Karsten Huffstadt

## Research Interests

- AI in Higher Education
- Digital Study Companions
- Human-AI Interaction & User Trust
- Technology Acceptance Model (TAM)

## Ongoing PhD Dissertation

"Optimizing Academic Studies Through AI and Digitalization:  
Conception, Development and Implementation of a Digital Study Companion"

## Publication Context

- Cumulative dissertation (7 papers)
- • Focus: Trust in auditive ChatGPT communication
- Conference: ACHI 2026 – IARIA

[marvin.tessitore.doktoranden@thws.de](mailto:marvin.tessitore.doktoranden@thws.de)

# Agenda

---

01 Introduction & Motivation

02 Research Gap & Objectives

03 Related Work

04 Theoretical Framework (TAM)

05 Methodology & Experimental Setup

06 Questionnaire Design

07 Results & Statistical Analysis

08 Discussion

09 Conclusion & Future Work

# Introduction & Motivation

The rise of conversational AI and the importance of user trust

## The AI Revolution

AI has become central to everyday digital interaction (OECD, 2019)

LLMs like ChatGPT generate coherent, human-like responses

Voice assistants (Alexa, Siri, Google) are widely adopted

Understanding human factors for acceptance is critical

## Why Trust Matters

Trust is a key determinant of successful Human-AI interaction

Shaped by: competence, transparency, contextual relevance

Prior attitudes toward AI influence trust formation

ChatGPT reliability varies across task types and domains

## Research Context

- ✓ 700M+ weekly active ChatGPT users (2025)
  - ✓ Voice-based interaction increasingly prevalent
  - ✓ Trust formation remains insufficiently understood
  - ✓ Existing research focuses on text-based interfaces
- ⚠ Open question: Does auditory interaction enhance or diminish perceived trustworthiness?

## Research Gap

- ▶ Impact of auditory interaction on trust in advanced LLMs remains insufficiently explored
- ▶ Voice interfaces offer natural communication, yet unclear if they enhance trust dynamics
- ▶ Text-based interfaces have been the primary focus of existing trust research

## Objective 1

Determine whether voice-based interaction influences trust differently than text-based interaction with ChatGPT

## Objective 2

Provide empirical evidence to support the design of transparent, reliable, and user-centered conversational AI systems

## Hypotheses

H<sub>0</sub>: Speech-based I/O has no effect on the degree of trust users place in ChatGPT.

H<sub>1</sub>: Speech-based I/O has an effect on the degree of trust users place in ChatGPT.

# Related Work

## Trust formation in digital environments and human-robot collaboration

### Trust and Risk in a Digital World

Heckersbruch et al. (2013)

- ▶ Face-to-face interaction restores trust compromised in electronic communication
- ▶ Interface design (aesthetics, usability, clarity) significantly affects credibility perceptions
- ▶ Transparency and perceived reliability influence trust in digitally mediated contexts
- ▶ Findings transferable to AI-driven systems operating within digital interfaces

### Improved Trust in Human-Robot Collaboration with ChatGPT

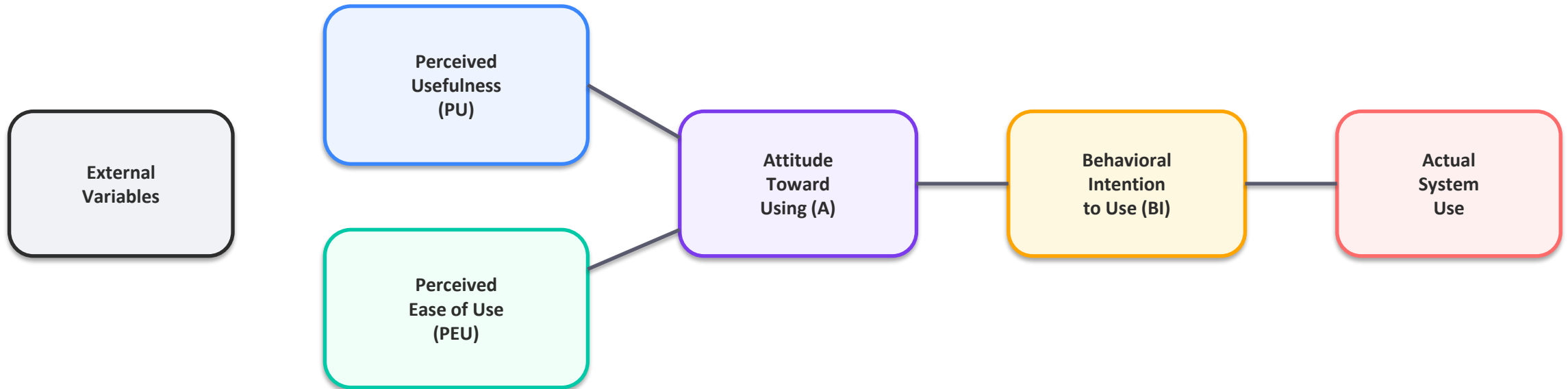
Ye, You & Du (2023)

- ▶ ChatGPT as conversational control interface in human-robot collaboration (HRC)
- ▶ 15 participants in VR environment
- ▶ Significant improvement in task efficiency vs. conventional fixed-command systems
- ▶ Reduced cognitive load, increased trust
- ▶ Natural language capabilities enhance trust in technologically mediated environments

**Key Insight:** Trust in AI systems is shaped by interface design, conversational competence, and interaction quality — not solely by content accuracy.

# Theoretical Framework

Technology Acceptance Model (TAM) — Davis (1989)



## Perceived Usefulness (PU)

The degree to which a person believes that using ChatGPT enhances their performance or supports task accomplishment.

## Perceived Ease of Use (PEU)

The extent to which a person believes that interacting with ChatGPT requires minimal effort and is intuitive to use.

# Methodology & Experimental Setup

Controlled A/B experiment with two interaction modalities

## Test A: Auditory Interaction

- ▶ Custom prototype: Laptop + ChatGPT + Azure Speech Services + C++
- ▶ Spoken input → text (Azure STT) → ChatGPT → speech output (TTS)
- ▶ Continuous listening mode (no manual activation)
- ▶ Quiet office environment, individual sessions
- ▶ Story-based tasks requiring nuanced reasoning

## Test B: Text-Based Interaction

- ▶ Standard ChatGPT web interface
- ▶ Keyboard input → written responses
- ▶ Continuous text-based dialogue
- ▶ Controlled, distraction-free environment
- ▶ Identical task description as Test A

## Participants & Procedure

**N = 66**

Participants

**33 / 33**

Per Group

**Ø 23.51**

Mean Age (yrs)

**5–7 min**

Session Duration

# Questionnaire Design

9-item TAM-based questionnaire | 5-point Likert scale

Item	Question	TAM Dimension
Q1	Do you generally trust machines or systems?	Trust
Q2	Do you think you have acquired new knowledge?	Usefulness
Q3	Would you pass on what you learned to others?	Usefulness (private)
Q4	Would you use the information for academic work?	Usefulness (research)
Q5	Were your questions answered competently?	Usefulness
Q6	Did you feel sufficiently supported by the system?	Ease of Use
Q7	Did you doubt the information provided?	Trust
Q8	How truthful do you consider the generated text?	Trust
Q9	Would you use the system again for learning?	Ease of Use & Usefulness

# Results: Statistical Analysis

Independent two-sample t-test | IBM SPSS Statistics

Item	Description	p-value	Significant?
Q1	General trust in machines	p = 0.019	✓ Yes (p = 0.019, d = 0.59)
Q2	Acquired new knowledge	p = 1.000	✗ No
Q3	Pass on learned content	p = 0.131	✗ No (trend, p = 0.131)
Q4	Use for academic work	p = 0.066	✗ No (trend, p = 0.066)
Q5	Questions answered competently	p = 0.615	✗ No
Q6	Sufficiently supported by system	p = 0.171	✗ No
Q7	Doubted system information	p = 0.082	✗ No
Q8	Truthfulness of generated text	p = 1.000	✗ No
Q9	Use system again for learning	p = 0.515	✗ No

## Summary

**1 of 9 items**

show significant differences

### Significant:

Q1 only (General Trust in Machines)

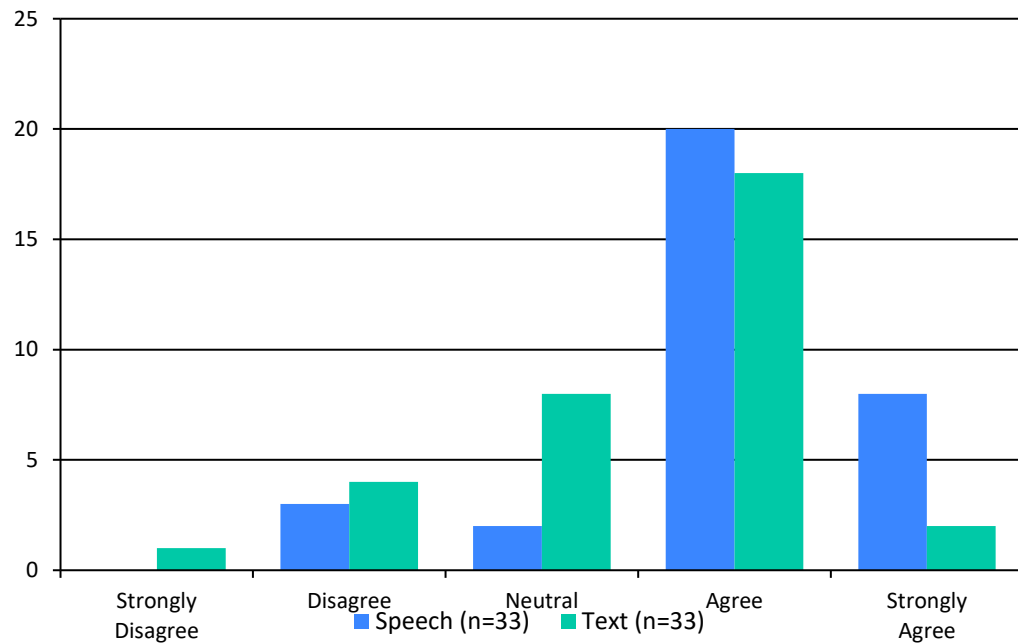
### Trend (†) / Not significant:

Q2, Q3†, Q4†, Q5, Q6, Q7, Q8, Q9

Reliability: Cronbach's  $\alpha = 0.547$  — Moderate consistency, reflecting the multi-construct design (Trust + Usefulness + Ease of Use) rather than methodological inadequacy.

# Results: Trust

Q1 (General Trust) | Q7 (Doubts) | Q8 (Truthfulness)



Q1: General Trust in Machines/Systems

## Trust Category Findings

### Q1: General Trust — $p = 0.019$ ✓

Speech group reported significantly higher general trust ( $M = 4.00$  vs.  $M = 3.49$ ;  $t(64) = 2.41$ ,  $p = 0.019$ ,  $d = 0.59$ ). Auditory modality activates a broader, dispositional sense of confidence in the technology.

### Q7: Doubts about Information — $p = 0.082$ ✗ ( $d = 0.44$ )

Approaching significance ( $p = 0.082$ ,  $d = 0.44$ ). Speech group expressed slightly more doubt ( $M = 2.18$ ) vs. text group ( $M = 1.79$ ). Not conclusive — warrants larger-sample replication.

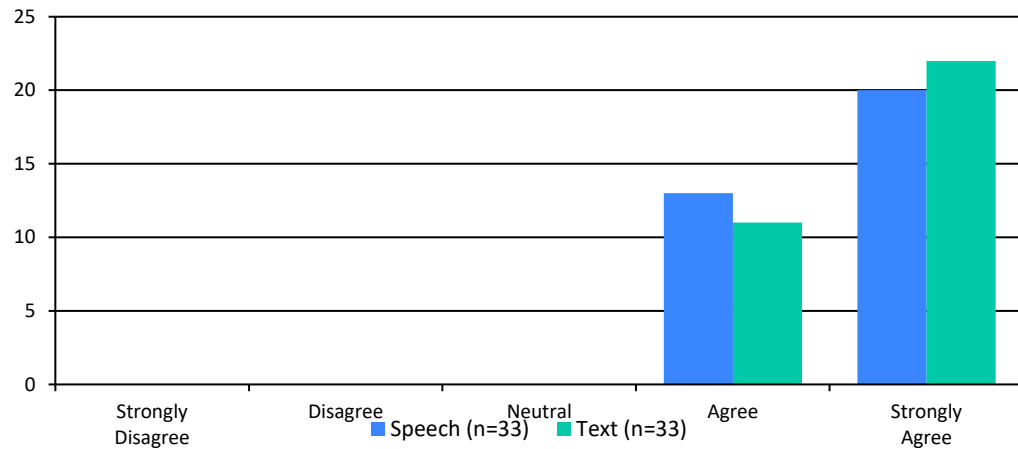
### Q8: Perceived Truthfulness — $p = 1.000$ ✗ ( $d = 0.00$ )

No difference whatsoever ( $p = 1.000$ ,  $d = 0.00$ ). Modality has zero effect on perceived truthfulness of ChatGPT outputs.

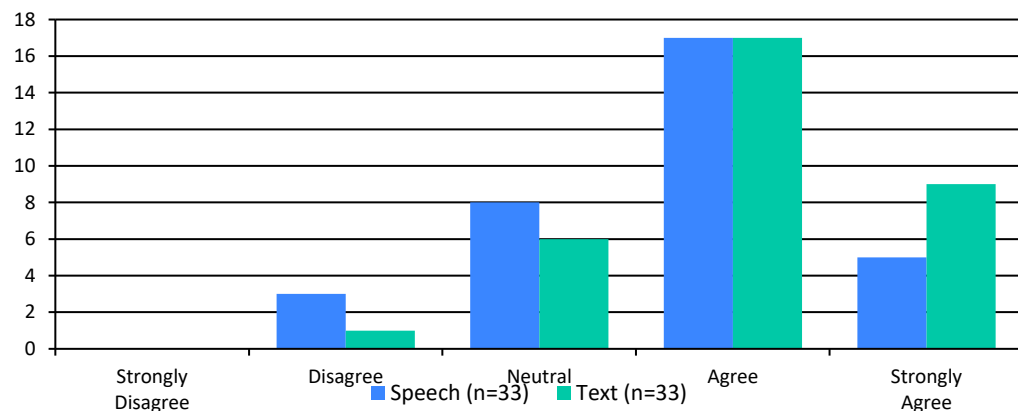
→ Speech significantly enhances generalized trust (Q1 only). It does NOT make users evaluate ChatGPT content as more truthful or credible.

# Results: Usefulness

Q2 (Knowledge) | Q3 (Share) | Q4 (Academic) | Q5 (Competence) | Q9 (Reuse)



Q5: Questions Answered Competently



Q3: Would Pass On Learned Content

## Usefulness Category Findings

**Q2: Knowledge Acquisition** —  $p = 1.000$  ✗

Both modalities equally effective ( $p = 1.000$ ,  $d = 0.00$ )

**Q3: Information Sharing** —  $p = 0.131$  ✓

Text group slightly higher ( $M_{\text{text}} = 4.03$  vs.  $M_{\text{speech}} = 3.73$ ) — directional trend, NOT significant

**Q4: Academic Reuse** —  $p = 0.066$  ✓

Text group slightly higher ( $M_{\text{text}} = 3.82$  vs.  $M_{\text{speech}} = 3.27$ ) — directional trend, NOT significant

**Q5: Response Competence** —  $p = 0.615$  ✓

No difference. Both groups rated response competence highly ( $M_{\text{speech}} = 4.61$ ,  $M_{\text{text}} = 4.67$ )

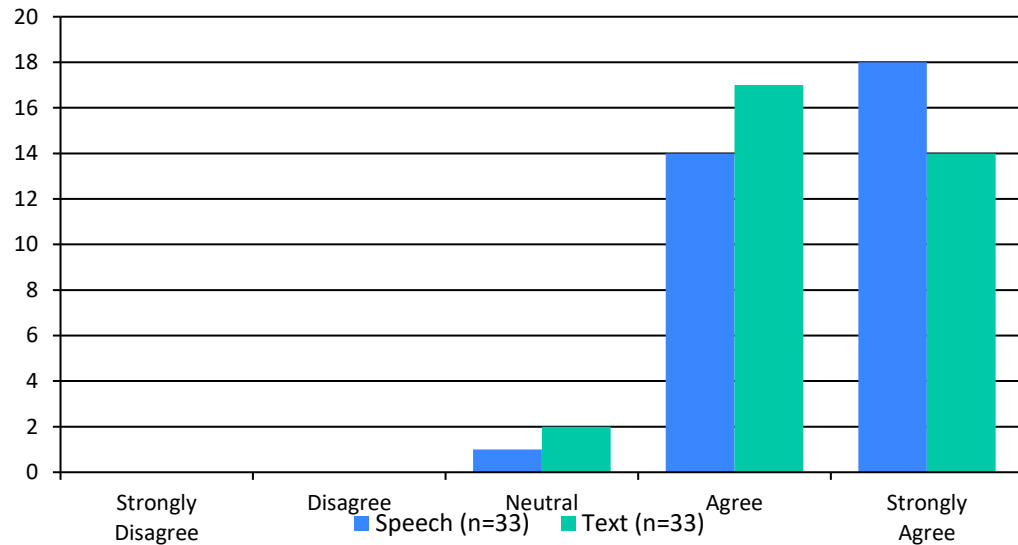
**Q9: System Reuse** —  $p = 0.515$  ✗

Comparable willingness across both groups ( $p = 0.515$ ,  $d = 0.16$ )

→ No significant differences in usefulness. Directional trends for Q3/Q4 (text slightly higher) require replication with larger, more diverse samples.

# Results: Ease of Use

Q6 (System Support) | Q9 (Willingness to Reuse)



Q6: Sufficiently Supported by System

## Ease of Use Findings

**Q6: System Support** —  $p = 0.171$  ✗

No significant difference. Both modalities perceived as equally supportive.

**Q9: Willingness to Reuse** —  $p = 0.515$  ✗

No significant difference. Comparable engagement across both groups.

### Key Finding:

Speech-based interaction does **NOT** introduce additional usability barriers despite lower familiarity for most users.

Both interaction modes were perceived as equally accessible and user-friendly.  
Voice-based AI does not face usability disadvantages compared to text-based interfaces.

# Discussion: Integrated Summary

## Trust

**Speech → Q1 only significantly higher**

- ▶ Speech: higher baseline trust in technology
- ▶ Q7 approaches ( $p=0.082$ ) but not significant; Q8  $p=1.000$
- ▶ Doubts and truthfulness ratings unchanged
- ▶ Modality influences broad attitudes, not granular judgments

## Usefulness

**Non-significant → Broadly equivalent**

- ▶ Both modalities: effective knowledge acquisition
- ▶ Text: directional trend for Q3/Q4 (not significant)
- ▶ No modality advantage for perceived competence (Q5  $p=0.615$ )
- ▶ Each modality has distinct advantages depending on user goals

## Ease of Use

**Equal → No meaningful differences**

- ▶ Both modes equally accessible
- ▶ Speech does not add usability barriers
- ▶ Comparable willingness to reuse
- ▶ Voice interaction is ready for mainstream adoption

## Key Takeaway

---

Trust formation depends less on the interaction channel per se and more on underlying system qualities such as accuracy, coherence, and conversational competence.

### **Speech**

Enhances generalized trust and perceived competence

### **Text**

Supports information sharing and academic reusability

### **Both**

Equally easy to use with comparable reuse intention

# Practical Implications

---

Recommendations for designers of AI-driven voice systems

1

## Do not assume inherent trust benefits

Voice-based interfaces do not automatically improve user trust. Trust is driven by system quality, not modality.

---

2

## Prioritize response quality & accuracy

Resources should be directed toward enhancing factual accuracy, coherence, and transparent system behavior.

---

3

## Leverage modality-specific strengths

Use speech for perceived competence and immediacy. Complement with text for information retention and reusability.

---

4

## Design multimodal systems

The most effective strategy combines voice interaction (competence) with text options (precision & permanence).

---

# Limitations & Future Work

## Limitations

- ▶ Moderate sample size (N = 66)
- ▶ Homogeneous sample (mainly THWS students)
- ▶ Cronbach's  $\alpha = 0.547$  (multi-construct design)
- ▶ Cross-sectional design (single interaction)
- ▶ No control for individual differences (personality, AI experience)
- ▶ Original TAM version (no extended constructs)

## Future Research Directions

- ▶ Larger and more diverse participant samples
- ▶ Longitudinal designs (trust over time)
- ▶ Validated multi-dimensional trust scales
- ▶ Individual differences (personality traits, domain expertise, AI experience)
- ▶ Extended TAM versions (UTAUT)
- ▶ Qualitative methods for richer insights
- ▶ Multimodal interaction combinations

# Selected References

---

- [1] OECD (2019). Artificial Intelligence in Society. Paris: OECD Publishing.
- [2] Brown et al. (2020). Language Models are Few-Shot Learners. arXiv.
- [3] Hoff & Bashir (2015). Trust in Automation. *Human Factors*, 57(3), 407–434.
- [4] Venkatesh et al. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, 27(3), 425–478.
- [5] Mayer, Davis & Schoorman (1995). Integrative Model of Organizational Trust. *AMR*, 20(3).
- [6] Nass & Brave (2005). *Wired for Speech*. MIT Press.
- [13] Ng & Zhang (2025). Trust in AI chatbots: A systematic review. *Telematics & Informatics*.
- [14] Heckersbruch et al. (2013). Vertrauen und Risiko in einer digitalen Welt.
- [15] Ye, You & Du (2023). Improved Trust in Human-Robot Collaboration with ChatGPT. arXiv.
- [20] Davis (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance. *MIS Quarterly*.
- [22] Shen et al. (2023). In ChatGPT We Trust? arXiv.

**Full reference list available in the paper.**

# Thank You

## Questions & Discussion

---

M. Sc. Marvin Tessitore — [Marvin.tessitore.doktoranden@thws.de](mailto:Marvin.tessitore.doktoranden@thws.de)

M. Sc. Hakan Arda — [Hakan.Arda.doktoranden@thws.de](mailto:Hakan.Arda.doktoranden@thws.de)

Prof. Dr. habil. Nicholas Mueller — [Nicholas.mueller@thws.de](mailto:Nicholas.mueller@thws.de)

Prof. Dr. Karsten Huffstadt — [Karsten.huffstadt@thws.de](mailto:Karsten.huffstadt@thws.de)

Technical University of Applied Sciences Würzburg-Schweinfurt (THWS)

Faculty of Computer Science and Business Informatics