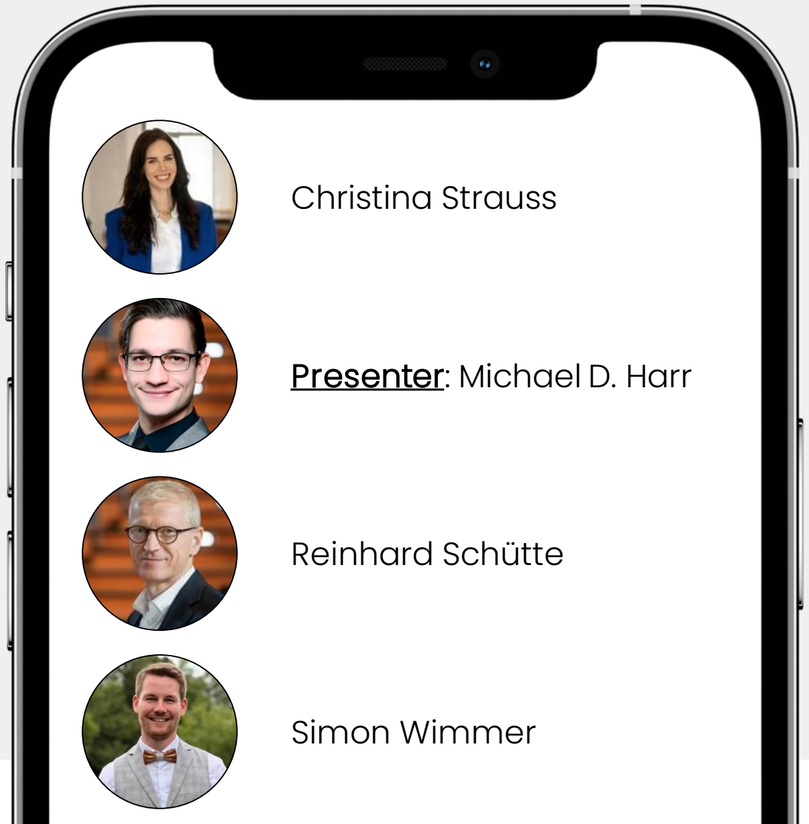
“Hey Siri, Don’t Make Me Mad” - Overcoming User Annoyances with Voice Assistants

Short Paper



UNIVERSITÄT DUISBURG ESSEN

European Conference on Information Systems (2024)

June 18th, 2024
Paphos, Cyprus

UNIVERSITÄT DUISBURG ESSEN

Motivation & Research Goal

Motivation

- **Substantial progress** in artificial intelligence (AI) & Large Language Models (LLMs) promises significant improvements in voice assistants (VAs) [8]
- Persisting user annoyances (UAs) remain a major challenge, undermining the potential for seamless human-like interaction

Research Gap

- Current research extensively covers the **technical capabilities** and **user experience design of voice assistants** but lacks a detailed exploration of how more sophisticated LLMs can specifically address and mitigate **user annoyances**.
- The distinction between advancements resulting from technical enhancements and those arising from enhanced human-like interaction qualities remains unanswered

Research Question

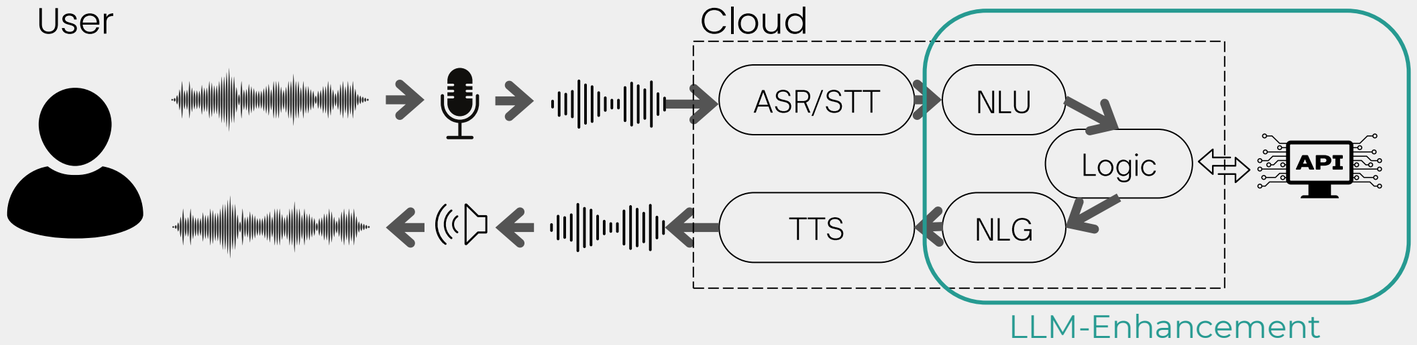
“To what extent do technically advanced and human-like LLMs mitigate UAs concerning the quality, conversational design, and comprehensibility of responses in VA interactions?”

UNIVERSITÄT DUISBURG ESSEN

Background

Voice Assistants & LLMs

- *VAs are manifestations of AI designed to interact in audible natural language [5,6]*
- **Operational paradigm**: Activation, Speech Input, Speech Processing, Execution, and Speech Output [4]



- **Added value**: non-domain-specific alignment, product smartness, humanness

User Annoyances

*UAs represent **significant impediments** to user satisfaction and engagement in the context of VAs [2]*

- **Implications**: User frustration, avoidance, and disengagement [9]
- **Sources**: Technology limitations, restrictive interaction designs, and unmet user expectations
- “Perceived humanness may be more important for users [...] than system response accuracy” [7, p. 64]

UNIVERSITÄT DUISBURG ESSEN

Method

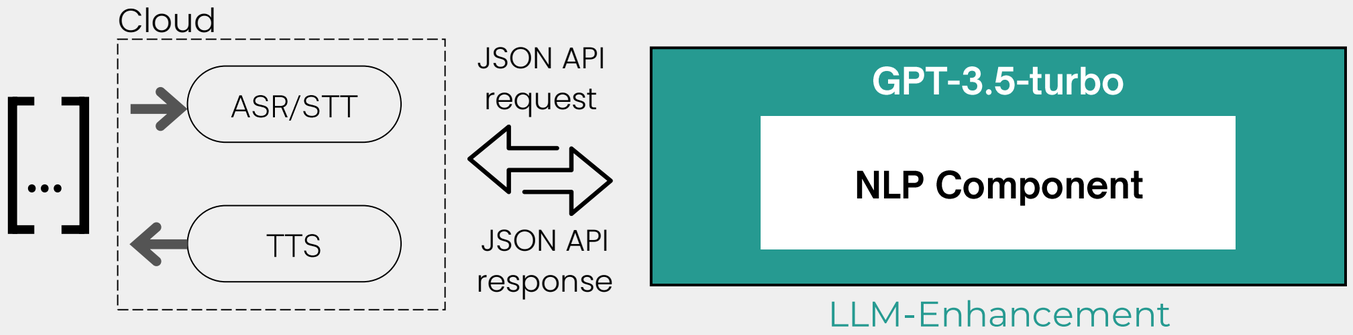
1) Baseline-survey on UAs
2) Development of "*Siri Pro*"
3) Semi-structured Interviews to assess UA mitigation

1)

- **Online-survey** with German-speaking respondents from two German universities via SurveyMonkey
- Questions encompassed **eight UAs** by Demaeght et al. (2022) [2], that can be influenced by the integration of an LLM (i.e., the NLP component)
- **21 Respondents** (female: 7; male: 14) who were familiar with Siri

2)

- **Development of "Siri Pro"** via "GPT-3.5-turbo" through an Apple Shortcut



3)

- **Semi-structured interviews after experimentation with "Siri Pro" [1]**
- Participants were asked the same questions as in 1) in a post-experiment survey but further interviewed on their impressions of the prototype and individual concerns
- **23 Respondents** (female: 5; male: 18) who were familiar with Siri

UNIVERSITÄT DUISBURG ESSEN

Preliminary Results

Baseline-Survey & Post-Experiment Survey

User Perceptions of Siri and "Siri Pro"	Preliminary Study (n=21)		Post-Experiment Survey (n=23)	
Question	**Mean**	**SD**	**Mean**	**SD**
When interacting with Siri/Siri Pro, did you have the feeling that...				
… Siri/Siri Pro has understood you in terms of content?	4.14	1.06	5.57	0.59
… Siri/Siri Pro remembers information that has already been given and you didn't have to repeat it.	3.95	1.43	5.61	0.66
… Siri's/Siri Pro's statement or question is understandable.	5.00	0.71	5.22	0.67
… Siri/Siri Pro's rarely repeats itself.	4.29	1.10	5.13	0.63
… Siri's/Siri Pro's answer is satisfactory.	3.95	0.97	4.87	0.81
… the results of your question are well presented.	3.95	1.20	4.09	1.16
… Siri's/Siri Pro's responses are accurate enough.	3.86	1.01	5.39	0.72
… there was a possibility of correction without having to start all over again.	2.62	1.32	5.30	1.10

Note. 6-point likert-scale: (1 – strongly disagree; 6 – strongly agree); SD = standard deviation

In-Depth Interviews

Most users based their assessments of "Siri Pro" on technical aspects; greater humanness was only rarely mentioned explicitly.

- **Response accuracy concerns**: Despite general confidence in the prototype's accuracy, about 10% of responses contained errors, such as incorrect facts and location directions
- **Privacy & Manipulation concerns**: Many participants expressed concerns about data privacy and potential manipulation of responses in voice assistants with integrated LLMs

UNIVERSITÄT DUISBURG ESSEN

Discussion & Outlook

Contribution & Discussion

We demonstrate, that the implementation of a refined LLM in the Siri VA can mitigate UAs.

- **Enhancements in Understanding and Context Retention**: Significant improvements in understanding user intent and preserving conversation context. This is attributed to LLMs' ability to reference past interactions and their extensive training on diverse datasets, which also contributed to reducing repetitive interactions

```
{"role": "user", "content": "Vorteile von React versus Angela"}

{"role":"assistant", "content":"React und Angular sind beide beliebte
JavaScript-Frameworks für die Entwicklung von Webanwendungen. [...]"}
```

- **Balanced Focus on Technical and Human Elements**: Technical advancements in VAs are crucial, but integrating human-like qualities could enhance user satisfaction. This highlights the need for VAs that combine technical proficiency with the ability to engage users empathetically [7]
- **Implications**: We demonstrate that integrating advanced LLMs allows VAs to transcend basic command-response roles, becoming dynamic communication partners

Outlook

- Integration of **more advanced LLMs** with real-time Internet access for Information retrieval + fine-tuning via prompt engineering and integrating social cues
- Further **focusing on user satisfaction and the distinguishment of both, technical advances and increased humanness** and their complex relation towards mitigating user annoyances and increasing user satisfaction

UNIVERSITÄT
DUISBURG
ESSEN

Thank You
for Your Attention!

UNIVERSITÄT
DUISBURG
ESSEN

Sources

[1] **Adeoye-Olatunde, O. A., and Olenik, N. L.** (2021). Research and scholarly methods: Semi-structured interviews. *Journal of the American College of Clinical Pharmacy*, *4*(10), 1358–1367.

[2] **Demaeght, A., Nerb, J., and Müller, A.** (2022). A survey-based study to identify user annoyances of German voice assistant users. In: F. Fui-Hoon Nah & K. Siau (Eds.), *HCI in Business, Government and Organizations. Lecture Notes in Computer Science* (pp. 261–271). Cham: Springer Gabler.

[3] **Guzman, A. L.** (2020). Ontological boundaries between humans and computers and the implications for Human-Machine Communication. *Human-Machine Communication*, 1, 37-54. https://doi.org/10.30658/hmc.1.3

[4] **Krol, B., and Boßow-Thies, S.** (2020). Akzeptanz von Sprachassistenten zur Steuerung von Smart Home Services. In: R. Buchkremer, T. Heupel, & O. Koch (Eds.), Künstliche Intelligenz in Wirtschaft & Gesellschaft (pp. 517–541). Wiesbaden: Springer Gabler.

[5] **Kulkarni, P., Mahabaleshwarkar, A., Kulkarni, M., Sirsikar, N., and Gadgil, K.** (2019). Conversational AI: An overview of methodologies, applications and future scope. 2019 5th International Conference On Computing, Communication, Control And Automation (ICCUBEA), 1–7. Pune, India: IEEE.

[6] **Kusal, S., Patil, S., Choudrie, J., Kotecha, K., Mishra, S., and Abraham, A.** (2022). AI-based conversational agents: A scoping review from technologies to future directions. *IEEE Access*, 10, 92337–92356.

[7] **Oesterreich, T. D., Anton, E., Schuir, J., Brehm, A., and Teuteberg, F.** (2023). How can I help you? Design principles for task-oriented speech dialog systems in customer service. *Information Systems and E-Business Management*, *21*(1), 37–79.

[8] **Schöbel, S., Schmitt, A., Benner, D., Saqr, M., Janson, A., and Leimeister, J. M.** (2023). Charting the Evolution and Future of Conversational Agents: A Research Agenda Along Five Waves and New Frontiers. *Information Systems Frontiers*.

[9] **Sharma, A., Dwivedi, R., Mariani, M. M., and Islam, T.** (2022). Investigating the effect of advertising irritation on digital advertising effectiveness: A moderated mediation model. *Technological Forecasting and Social Change*, *180*(May), 121731.

[10] **Turkle, S.** (1984). *The Second Self: Computers and the Human Spirit*, 1 ed. New York: Touchstone.