



In an era where artificial intelligence is fundamentally changing our perception of reality, the VR-experience *A Reverse Turing Test* presents a fascinating reversal of the classic Turing Test. This immersive Virtual Reality (VR) installation challenges visitors to disguise themselves as the only human among advanced AI systems, raising profound questions about the nature of intelligence, identity and reality.

## Description

The artwork *A Reverse Turing Test* is an interactive Virtual Reality (VR) installation that explores the boundaries between artificial and human intelligence, raising fundamental questions about human nature and reality in an increasingly digitalized world. At a time when AI systems and virtual humans are becoming increasingly sophisticated (Bubeck et al. 2023; Jones and Bergen 2024), our installation challenges participants to question their notions of authenticity, identity and human behaviour.

In our VR-Experience, human visitors find themselves in a virtual train compartment, embodying the persona of Genghis Khan as an avatar. The other passengers are famous personalities as well: Aristotle, Mozart, Leonardo da Vinci, and Cleopatra. However, they are driven by some of the most sophisticated AIs currently available: GPT4, Claude,

### Tore Knabe

Berlin, Germany,  
tamulur@yahoo.com

### Jonathan Harth

University Witten/Herdecke, Witten, Germany,  
jonathan.harth@uni-wh.de

**Keywords** Artificial Intelligence, Virtual Reality, Large Language Models, Human-Agent-Interaction, Turing Test

**DOI** [10.34626/2025\\_xcoax\\_028](https://doi.org/10.34626/2025_xcoax_028)

Llama, and Gemini. The train conductor informs the group that one of them is a human and they have to find out who. Aristotle suggests that each one asks the person to their left a question that helps find out whether the one answering is a human or AI. Thus, the installation reverses the classic Turing test (Turing, 1950): instead of a human trying to identify a machine, the AIs have to unmask the human among them.

This reversal creates a fascinating dynamic in which the human participant is forced to reflect on and adapt their behaviour in order to ‘pass’ as an AI. This game of identity, imitation and projection raises profound questions:

How do we define and recognise human behaviour in a world where AIs are becoming increasingly human-like?

To what extent does our behaviour change in environments that are increasingly populated by artificial intelligences?

The installation invites visitors to explore the boundaries between human and machine, past and future, reality and simulation, while asking fundamental questions about the nature of our increasingly digitalized existence (Sejnowski 2023).

## Technical Background

The experience was developed using the Unity Engine and runs on all headsets via OpenXR. At the start, each of the different characters is randomly assigned an LLM (for example, Aristotle is controlled by GPT-4o, Mozart by Claude Sonnet, etc.). Depending on who is currently up for questions/answers/voting, the corresponding LLM is prompted, and the answer from the respective character is then output via text-to-speech (Google and Azure) with real-time lip sync.

The prompt contains a description of the situation, the complete conversation so far, and the instruction. In the output, in addition to the spoken part, meta-information is provided in JSON—such as whether the current statement is an answer, a question to the next person, a vote, and if it is a vote, for whom, etc. The player’s voice is transcribed into text by speech-to-text services, and then, along with the prompts, sent to the respective LLM. The intro and outro are scripted; the LLMs only come into play when Aristotle poses the first question to Mozart.

Whilst we already have significant video and image material from the developer version, the installation is currently being rebuilt into a fully interactive version. This version will allow visitors to seamlessly go through the entire experience from start to finish.



**Fig.1.** Several AI-agents having a chat during the VR-experience *A Reverse Turing Test*.



**Fig.2.** The flow of the experience can be viewed in this video:  
[https://www.youtube.com/watch?v=MxTWM9vT\\_o](https://www.youtube.com/watch?v=MxTWM9vT_o)

**Acknowledgments.** This work is part of the research project *Theater der erweiterten Realitäten* (Theater an der Ruhr, Mülheim; MIREVI, University of Applied Sciences Düsseldorf; Academy for Theatre and Digitality, Dortmund). This work was supported in part by a grant from the Ministry of Culture and Science of the State of North Rhine-Westphalia as part of the program NEUE WEGE in cooperation with the NRW KULTURsekretariat.

## References

- Bubeck, Sébastien, Varun Chandrasekaran, Ronen Eldan, et al.** 2023. "Sparks of Artificial General Intelligence: Early Experiments with GPT-4". arXiv preprint, arXiv:2303.12712.
- Sejnowski, Terrence.** 2023. "Large Language Models and the Reverse Turing Test". *Neural Computation*, 35 (3): 309–342. [https://doi.org/10.1162/neco\\_a\\_01563](https://doi.org/10.1162/neco_a_01563).
- Jones, Cameron, and Benjamin Bergen.** 2024. "Does GPT-4 pass the Turing test?" arXiv preprint, arXiv:2310.20216.
- Turing, Alan.** 1950. "Computing Machinery and Intelligence." *Mind* 59: 433–460.