

## On the Way towards Self-aware AI?

New Technology Foresight Project Addresses the Vision of “Conscious” Machines



*High need for clarification: it is unclear and disputed whether machines may turn into conscious creatures thanks to artificial intelligence. (Image: Pixabay)*

**Alexa, Sophia, Watson: the ancient idea of a humanoid machine with superhuman powers has received fresh impetus by the progress achieved in AI research. Some actors offer the prospect of a “superintelligence” with self-awareness. But how realistic is this? Within the project “Clarification of Suspicion of Consciousness in Artificial Intelligence” funded by the Federal Ministry of Education and Research, technology assessment experts of Karlsruhe Institute of Technology (KIT) analyze this issue that has hardly been studied before.**

When the robot “Sophia” stepped up to the speaker’s desk at a conference in Riyadh, Saudi Arabia, and explained to the half amused, half astonished audience its self-image of a learning and communicating machine in human form, public perception considered this a milestone on the apparently ever shorter way to an “awakening” artificial intelligence reflecting its individuality and its inner states. Also

**Monika Landgraf**  
Chief Press Officer,  
Head of Corp. Communications

Kaiserstraße 12  
76131 Karlsruhe, Germany  
Phone: +49 721 608-21105  
Email: [presse@kit.edu](mailto:presse@kit.edu)

**Press contact:**

Justus Hartlieb  
Press Officer  
Phone: +49 721 608-21155  
Email: [justus.hartlieb@kit.edu](mailto:justus.hartlieb@kit.edu)

AI-based systems, such as the smart speaker “Alexa,” IBM’s interoperable super brain “Watson” or Google’s self-learning chess giant “AlphaZero,” boost the vision of a “superintelligence” which will be feasible within a foreseeable time and overshadow anything ever in the world. Renowned actors in science and the arts warn of the impacts of such a change of epochs, others speak of sheer utopic opportunities. Surprisingly, it is hardly ever asked what “conscious AI” is and whether the scenarios of machines with an own existence may become reality.

This is the point of departure of the two-year research project “Clarification of Suspicion of Consciousness in Artificial Intelligence (AI Consciousness).” “We want to demystify the topic of AI consciousness,” says project leader Professor Karsten Wendland of KIT’s Institute for Technology Assessment and Systems Analysis (ITAS). The computer scientist and technology assessment expert continues: “some consider it impossible that machines, in particular AI systems, will ever become ‘conscious’. Others say that conscious AI systems already exist and still hide from us. We want to develop a common, transdisciplinary understanding and bring the results into public discourse.”

To first analyze the current status of AI consciousness, the project team will use a mixed methods approach, which consists of a systematic acquisition of the debates in the special disciplines and in public discourse, expert interviews, and a bibliometric media analysis. In this way, typical positions, arguments, and narratives will be identified and described for the first time ever. “As analysts,” project leader Wendland says, “we adopt a neutral attitude. At the same time, we will not shrink back from including in our study highly courageous positions on AI consciousness that are rather rare in the European region, for instance, that it is not needed to distinguish between humans and machines, as they both are the results of creation.” In a second step, experts, who have not yet been talking to each other, will be brought together to discuss central issues relating to AI consciousness. These transdisciplinary bridges, substantiated in workshops and symposiums, are to result in a transdisciplinary understanding of this both mysterious and delicate topic.

An important part of the project will be the generally understandable presentation of the state of knowledge developed by analysis and discourse and the communication of this knowledge to the interested public as well as to stakeholders and other scientific communities. This will be achieved by citizen dialog formats as well as by podcasts, brochures, and a planned website. In addition, close exchange with scientific editors is envisaged. In line with its demystifying and clarify-

ing approach, the project team wants to disseminate the results relating to the awakening AI consciousness in a transparent and still substantiated manner. “In this way, we want to extend and maybe even purify common knowledge and take up the suggestions resulting from this dialog process.”

The project “Clarification of Suspicion of Consciousness in Artificial Intelligence (AI Consciousness)” is scheduled for a duration of two years until December 2020. Within the funding program “Innovations- und Technikanalyse” (ITA, innovation and technology assessment), it is funded by the Federal Ministry of Education and Research with a total of EUR 335,000.

**More information:**

[https://www.itas.kit.edu/english/projects\\_wend19\\_kibew.php](https://www.itas.kit.edu/english/projects_wend19_kibew.php)

GEFÖRDERT VOM



Bundesministerium  
für Bildung  
und Forschung

**Being “The Research University in the Helmholtz Association,” KIT creates and imparts knowledge for the society and the environment. It is the objective to make significant contributions to the global challenges in the fields of energy, mobility and information. For this, about 9,300 employees cooperate in a broad range of disciplines in natural sciences, engineering sciences, economics, and the humanities and social sciences. KIT prepares its 25,100 students for responsible tasks in society, industry, and science by offering research-based study programs. Innovation efforts at KIT build a bridge between important scientific findings and their application for the benefit of society, economic prosperity, and the preservation of our natural basis of life.**

This press release is available on the internet at [http://www.sek.kit.edu/english/press\\_office.php](http://www.sek.kit.edu/english/press_office.php).

The photo in the best quality available to us may be downloaded under [www.kit.edu](http://www.kit.edu) or requested by mail to [presse@kit.edu](mailto:presse@kit.edu) or phone +49 721 608-21105. The photo may be used in the context given above exclusively.

This year's **anniversary logo** recalls the milestones reached by KIT and its long tradition in research, teaching, and innovation. On October 1, 2009, KIT was established by the merger of its two predecessor institutions: the Polytechnic School and later University of Karlsruhe was founded in 1825, the Nuclear Reactor Construction and Operation Company and later Karlsruhe Research Center in 1956.