

# “Voices on Conscious Machines – A Comprehensive Survey Analysis”

Rapid developments in Artificial Intelligence (AI) are increasingly blurring the boundaries between humans and machines. As the capabilities of AI systems advance, particularly of Large Language Models (LLM) such as ChatGPT and Gemini, questions regarding their intelligence level and potential to develop consciousness arise. Instances like Google engineer Blake Lemoine claiming that Google’s AI Lambda is sentient, or Facebook’s AI chatbots developing a language that could no longer be understood by its developers, highlight these concerns [5, 14]. Moreover, Sophia, a humanoid robot granted citizenship in Saudi Arabia and became the first non-human citizen, raising numerous legal, social, and moral considerations [2]. These rapid advancements and notable events have sparked debates about AI safety & ethics. The European Commission’s Artificial Intelligence Act, passed by the European Parliament in 2024, categorizes AI into four risk levels, reflecting a growing need for regulatory oversight [4]. The UN Secretary-General’s AI Advisory Body released: “Governing AI for Humanity”, which proposes to enhance international governance of AI to achieve the Sustainable Development Goals (SDG) [16]. More concerns are expressed by public figures like Elon Musk, warning AI could pose an existential threat to humanity, marking a pivotal point in history where humans create something more intelligent than themselves [7]. All these events raise questions regarding AI’s level of intelligence and potential to develop consciousness.

For decades, intelligence and consciousness have been explored in various fields, such as philosophy, neuroscience, and computer science. Although many theories have significantly shaped our understanding, there is no common ground of what exactly those are, because of their subjective and abstract nature [1, 3, 6, 10, 12, 13, 15]. Based on those theories we want to further explore if they also apply to Artificial Intelligence. Recent studies explore Conscious AI by targeting it from different angles and reveal that current AI Systems are perceived as conscious to some extent [8, 9, 11]. However, these studies primarily focus on public opinion.

To gain a deeper understanding, it is essential to consider the perspectives of experts. To address this gap, this master’s thesis will expand on previous studies by seeking insights from experts in neuroscience, computer science, and philosophy, as these fields are not only essential for understanding consciousness and intelligence but also influential in directing future developments. This research aims to identify the perceived level of consciousness and intelligence of modern and future technologies through a survey. By comparing these findings with the public perception, as well as different groups and research fields, it aims to provide valuable insights, understand the direction of AI development, and guide future research in the field.

The proposed work consists of the following parts:

- Literature research in intelligence, consciousness, machine learning, and conscious AI
- Develop and conduct a survey, targeting around 100 experts, including students, professors, and researchers in neuroscience, computer science, and philosophy
- Analyze, evaluate, and present results, including statistical differences between groups and fields, as well as thematic analysis for qualitative questions

The thesis must contain a detailed description of all developed and used algorithms as well as a profound result evaluation and discussion. The implemented code has to be documented and

provided. An extended research on literature, existing patents and related work in the corresponding areas has to be performed.

**Advisors:** Madeleine Flaucher M.Sc., Hamid Moradi M.Sc., Prof. Dr. Björn Eskofier, Dr.-Ing. Heike Leutheuser  
**Student:** Kevin Kremer  
**Start—End:** 15.07.2024—15.01.2025

## References

- [1] **Binet, A., & Simon, T. (1916c).** The development of intelligence in children (The Binet-Simon Scale). In Williams & Wilkins Co eBooks. <https://doi.org/10.1037/11069-000>
- [2] **British Council. (n.d.).** Should robots be citizens? <https://www.britishcouncil.org/anyone-anywhere/explore/digital-identities/robots-citizens> (accessed 28.05.2024)
- [3] **Chalmers, D. J. (1995).** Philosophy Program, Research School of Social Sciences, & Australian National University. Facing up to the problem of consciousness. In *Journal of Consciousness Studies* (Vols. 2–3, pp. 200–219). <https://consc.net/papers/facing.pdf>
- [4] **European Parliament. (2024).** Artificial Intelligence Act. In TEXTS ADOPTED. [https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138\\_EN.pdf](https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf)
- [5] **Field, M. (2018).** Facebook shuts down robots after they invent their own language. *The Telegraph*. <https://www.telegraph.co.uk/technology/2017/08/01/facebook-shuts-robots-invent-language/> (accessed 28.05.2024)
- [6] **Gardner, H. (1987).** The Theory of Multiple Intelligences. *Annals of Dyslexia*, 37, 19–35. <http://www.jstor.org/stable/23769277>
- [7] **Landi, M. (2023).** Elon Musk says AI one of the 'biggest threats' to humanity. *The Independent*. <https://www.independent.co.uk/news/uk/politics/elon-musk-rishi-sunak-government-spacex-tesla-b2439715.html> (accessed 28.05.2024)
- [8] **Lima, G. et al. (2020).** Collecting the public perception of AI and robot rights. *Proceedings of the ACM on Human-computer Interaction*, 4(CSCW2), 1–24. <https://doi.org/10.1145/3415206>
- [9] **Scott, A. E. et. al. (2023).** Do You Mind? User Perceptions of Machine Consciousness. <https://doi.org/10.1145/3544548.3581296>
- [10] **Searle JR. (1980).** Minds, brains, and programs. *Behavioral and Brain Sciences*;3(3):417-424. doi:10.1017/S0140525X00005756
- [11] **Spatola, N. (2019).** Switch off a Robot, Switch off a Mind? <https://doi.org/10.1145/3349537.3351897>
- [12] **Spearman, C. (1904b).** "General Intelligence," Objectively Determined and Measured. *the American Journal of Psychology/American Journal of Psychology*, 15(2), 201. <https://doi.org/10.2307/1412107>
- [13] **Sternberg, R. J. (1985).** *Beyond IQ: A triarchic theory of human intelligence*. Cambridge University Press.
- [14] **Tiku, N. (2023).** The Google engineer who thinks the company's AI has come to life. *Washington Post*. <https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/> (accessed 28.05.2024)
- [15] **Turing, A. M. (1950).** I.—COMPUTING MACHINERY AND INTELLIGENCE. *Mind*, LIX(236), 433–460. <https://doi.org/10.1093/mind/lix.236.433>
- [16] **UN. High-Level Advisory Body on AI. (2023).** Interim report: Governing AI for Humanity. [https://www.un.org/sites/un2.un.org/files/ai\\_advisory\\_body\\_interim\\_report.pdf](https://www.un.org/sites/un2.un.org/files/ai_advisory_body_interim_report.pdf)