

Towards explainable artificial intelligence for social applications: analysis of brain factors in the context of trust and attractiveness assessment

Abstract

Assessing trust and attractiveness is one of the most fundamental yet difficult social processes to describe. These assessments are so immediate and easy that we rarely think about the complex neural processes behind them. However, understanding these mechanisms is crucial, not only for cognitive science, but also for the design of artificial intelligence that can interact with humans in a meaningful way.

The article proposes the concept of Explainable AI, which combines EEG data with algorithms that learn to recognize and visualize neural patterns associated with the assessment of trust and attractiveness. The system is designed to translate brain processes into a form that is understandable to humans, bridging the gap between neurobiology and artificial intelligence. Its theoretical foundations, the limitations of existing approaches, and possible practical applications and ethical implications are presented.

The proposed solution is a step towards a technology that not only predicts human decisions, but also explains them in a way that is close to how humans themselves understand others.