

TOWARDS SOCIALLY INTERACTIVE AGENTS WITH EXPLANATORY SKILLS

Elisabeth André

*Universität Augsburg
andre@informatik.uni-augsburg.de*

Abstract: With the increased complexity of ML algorithms, sophisticated methods for explaining decisions made by an AI system are becoming an urgent need. They enable users to scrutinize a system's decisions and help users calibrate their trust towards the system and eventually make better use of it. In this talk, we explore the potential of socially interactive agents to communicate the rationale behind the decisions of an AI system to human users. Unlike typical explanatory components, such agents do not only provide audio-visual explanations, but also engage with the human user in a social interaction. This step also requires moving from the pure presentation of relevant information to multimodal explanations embedded in a narrative. In my talk, I will report on various user studies we conducted in order to investigate how different kinds of explanation provided by a socially interacting agent are perceived by users and to what extent they support them in building up mental model of the underlying AI system. The talk will be illustrated by examples from various international and national projects from the area of social coaching and health care.