Title: Reasoning about Trust

Speaker: Andreas Herzig (CNRS, IRIT France)

Venue: KD101

Date: 10 February 2025

Time: 3:30PM

Abstract:

Several informal definitions of the concept of trust exist in the social sciences literature. These definitions view trust as trust in a human being. The opacity of AI systems has motivated the study of trust in such systems. It also motivates the design of formal systems that allow us to reason about trust.

We are going to focus on a semi-formal definition of trust that had been proposed by Cristiano Castelfranchi and Rino Falcone in a series of papers and that was influential in particular in the multi-agent systems community. We are going to present a simple logic of belief and agency that is rich enough to account for the relevant concepts defining trust. While similar logics were proposed before, they are highly complex and only give little formal results. In contrast, we design a minimalist logic that provides what we claim to be the simplest possible logic of trust.