

Ophelia Deroy (LMU Munich), 01.06.22

How will humans treat AI? Or, Human social intelligence meets artificial intelligence

Much of our fears and regulations focus on anticipating how AI may treat human beings, and trying to avoid damaging scenarios. The same is true in AI ethics, which wants to make sure that AI agents remain transparent, trustworthy or benevolent. But how will humans behave if they face an all-benevolent AI? Isn't uncertainty about other's intentions part of what makes us more cautious, and ready to negotiate?

This talk will address these questions, and highlight the importance and challenges of extending theories of human cooperation and social cognition to us interacting with AI.

Short bio : Ophelia Deroy is the former co-director of the Institute of Philosophy in London, and holds the Chair for philosophy of Mind at the LMU, where she leads the research group "Cognition, Values and Behaviour" (CVBE). She is an expert in social cognition, and is particularly interested in non-linguistic interactions. With colleagues, she recently started a project on human-AI interactions ("Artificial partners" see <https://www.ai-partners.org>).