

Käte Hamburger Kolleg: Cultures of Research
Lecture Series 2022 “Philosophy of AI_Optimists and Pessimist Views”

Jose Hernandez Orallo (Universitat de Politècnica Valencia), 18.05.2022

Calibrating expectations about AI: A renewed endeavour towards the measurement of behaviour

A lecture series on optimistic and pessimistic views of AI can serve to calibrate our expectations. There are questions such as how intelligent AI systems really are, and will be, or how they are going to change the world, starting from our own cognition and the very concept of intelligence. One major way of addressing these questions is through measurement. Indeed, dealing successfully with AI and the future of cognition more broadly depends on a renewed perspective on the evaluation of behaviour in all its varieties. In this talk, I will introduce a series of endeavours about the present and future measurement of AI, such as the evaluation of capabilities rather than task performance, the evaluation of general-purpose systems rather than specialised ones, the evaluation of AI extenders rather than externalised systems, the evaluation of the transformative effect on skills in the workplace, etc. The sheer complexity of these challenges and the pace in which AI is developing suggests a pessimistic view on intelligence evaluation: the theory and tools for the measurement of behaviour would be doomed to lag behind the cognitive transformation in the decades ahead. However, to counterbalance this gloomy view, I will vindicate some key elements: the identification of the dimensions of difficulty to determine capability and generality profiles, the proper study of instance variation to ensure robustness in evaluation, the need of more ambitious meta-analyses of experimental data about AI measurement and the radical proposal that every deployed AI system should come with an assessor, an external model predicting and explaining its behaviour.