LEE KUAN YEW CENTRE FOR INNOVATIVE CITIES
PRESENTS

BROWNBAG LUNCHTIME TALK

In AI We Trust: How Humans Teach an AI to Reciprocate Trust

DATE AND TIME
30 July 2024,
1 - 2 pm

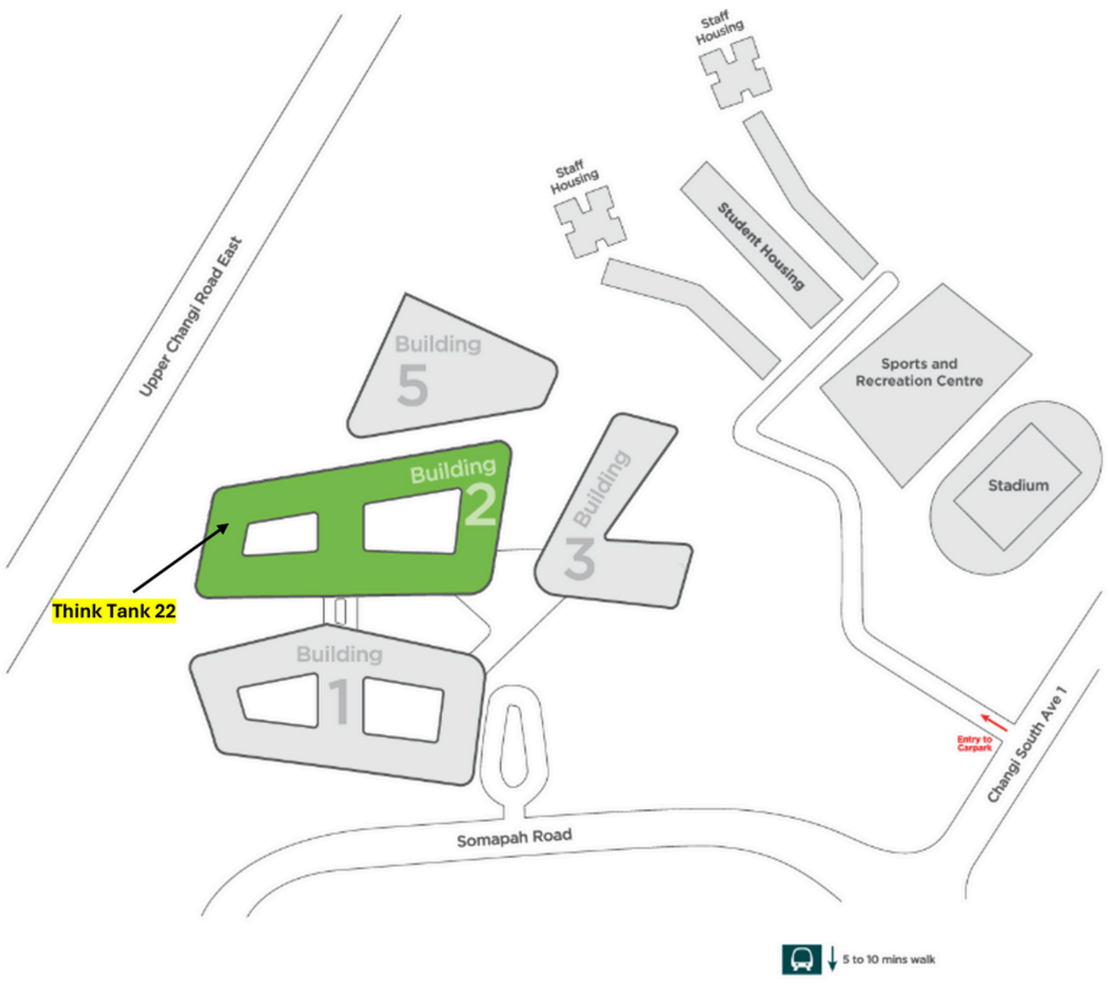
DR BENJAMIN PRISSE
Research Fellow,
Lee Kuan Yew Centre for Innovative Cities

VENUE
Think Tank 22
Building 2, Level 3 (2.311)

SYNOPSIS - Trust in Artificial Intelligence (AI) is a topic that has been on the rise since AI-enhanced technologies became common in a variety of domains. Trust-based human-AI interactions are becoming increasingly common in everyday life. However, little is known about whether we truly trust AI or what can be done for us to trust AI. This talk will present findings from a study exploring whether humans trust an AI when they are given the opportunity to teach the AI how to reciprocate the trust of a human being. With an AI initially possessing four strategies to reciprocate trust (minimum reciprocity, slight reward for trusting, fair sharing of benefits, and generosity), subjects were able to teach or unteach these strategies to the AI according to the outcomes in each period. Results indicate that subjects rapidly unteach the two least profitable strategies to the AI, then figure out the most profitable strategy.

DR BENJAMIN PRISSE
Research Fellow,
Lee Kuan Yew
Centre for
Innovative Cities

DR BENJAMIN PRISSE is a Research Fellow at the Lee Kuan Yew Centre for Innovative Cities (LKYCIC) at the Singapore University of Technology and Design (SUTD). He did a Master of Economics at Toulouse School Economics, then a Master of Neuroeconomics at Maastricht University. He recently completed a PhD in Data Science with a particular focus on Experimental Economics at University Loyola Andalucía. His research interests are visual experiments, online environments, Artificial Intelligence, experiments with teenagers.



Scan QR Code to Register

Lee Kuan Yew Centre for Innovative Cities

For more information on the event, please visit https://lkycic.sutd.edu.sg/

For enquiries, please email lkycic@sutd.edu.sg