



ISTA Lecture

# Provably Beneficial Artificial Intelligence

**Stuart J. Russell (Berkeley, University of California)**

**Host: Thomas A. Henzinger**

We are deeply pleased and proud to invite you to our third ISTA Lecture with our special and distinguished guest, [Prof. Dr. Stuart J. Russell OBE](#).

## Abstract

As AI advances in capabilities and moves into the real world, its potential to benefit humanity seems limitless. Yet we see serious problems including racial and gender bias, manipulation by social media, and an arms race in lethal autonomous weapons. Looking further ahead, Alan Turing predicted the eventual loss of human control over machines that exceed human capabilities. I will argue that Turing was right to express concern but wrong to think that doom is inevitable. Instead, we need to develop a new kind of AI that is provably beneficial to humans.

## About Stuart J. Russell

Stuart Russell is a Professor of Computer Science at the University of California at Berkeley, holder of the Smith-Zadeh Chair in Engineering, and Director of the Center for Human-Compatible AI. He is a recipient of the IJCAI Computers and Thought Award and held the Chaire Blaise Pascal in Paris. In 2021 he received the OBE from Her Majesty Queen Elizabeth and gave the Reith Lectures. He is an Honorary Fellow of Wadham College, Oxford, an Andrew Carnegie Fellow, and a Fellow of the American Association for Artificial Intelligence, the Association for Computing Machinery, and the ACM SIGACTS and SIGPLAN divisions of the ACM. His book "Artificial Intelligence: A Modern Approach" (with Peter Norvig) is the standard text in AI, used in 1500 universities in 135 countries. His research covers a wide range of topics in artificial intelligence, with a current emphasis on the long-term future of artificial intelligence and its relation to human values. He has developed a new global seismic monitoring system for the nuclear test ban treaty and is currently working to ban lethal autonomous weapons.



ISTA Lecture **Young Lounge**

Ahead of the lecture, our Young Lounge format offers high school students and their teachers the

possibility to interact with the speaker in a less formal setting. For details and registration please go to the [ISTA Lecture Young Lounge page](#).

### **How to get here**

Public transportation. By showing the invitation to this lecture, participants can use the ISTA Shuttle (Bus #142) from Heiligenstadt for free. Depart from the [Heiligenstadt bus stop](#) at 15:57 and arrive at 16:24. Please see the [shuttle timetable](#).

Private transportation. Parking options for the lecture will be communicated via email to registered participants.

### **REGISTRATION**

[Click here](#) or on the button below to register for the attendance in the Raiffeisen Lecture Hall (limited capacity!).