

20 BIAS AND TRUST IN AI

How can we come to trust that an AI system or application will give us good decision support and not produce a harmful action? In the end this amounts to doing a lot of testing, including stress testing, to find out how the system performs and whether it meets its specifications.

Many AI systems rely on large data sets for training and validation. When these data contain biases, the machine can acquire the biases during its learning and training. The bias can make the machine behave in unintended ways, undermining our trust. Engineers and statisticians have long struggled learn how to detect and remove biases from data. This quest has now come to AI system developers and data curators.

Professor Josh Kroll gives us an AI perspective on these questions.