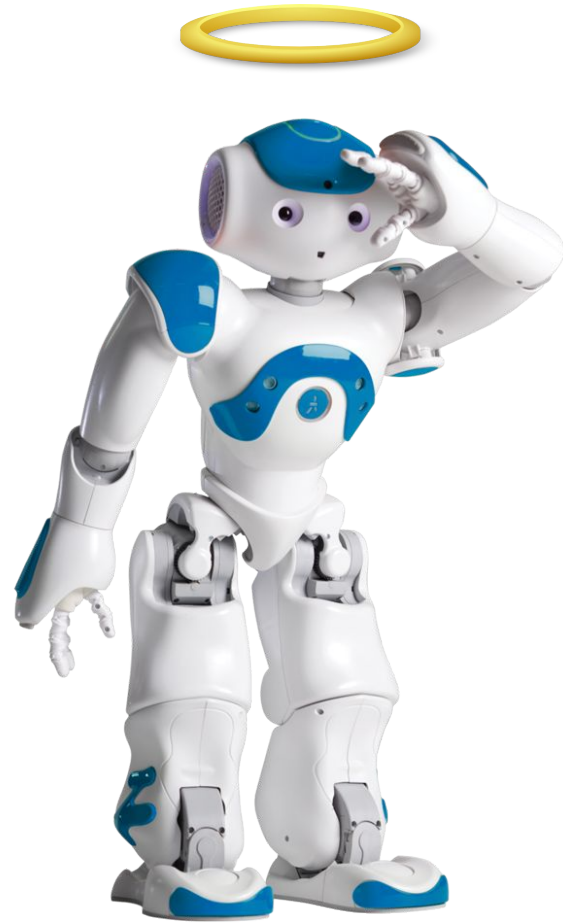


Ideas Lab

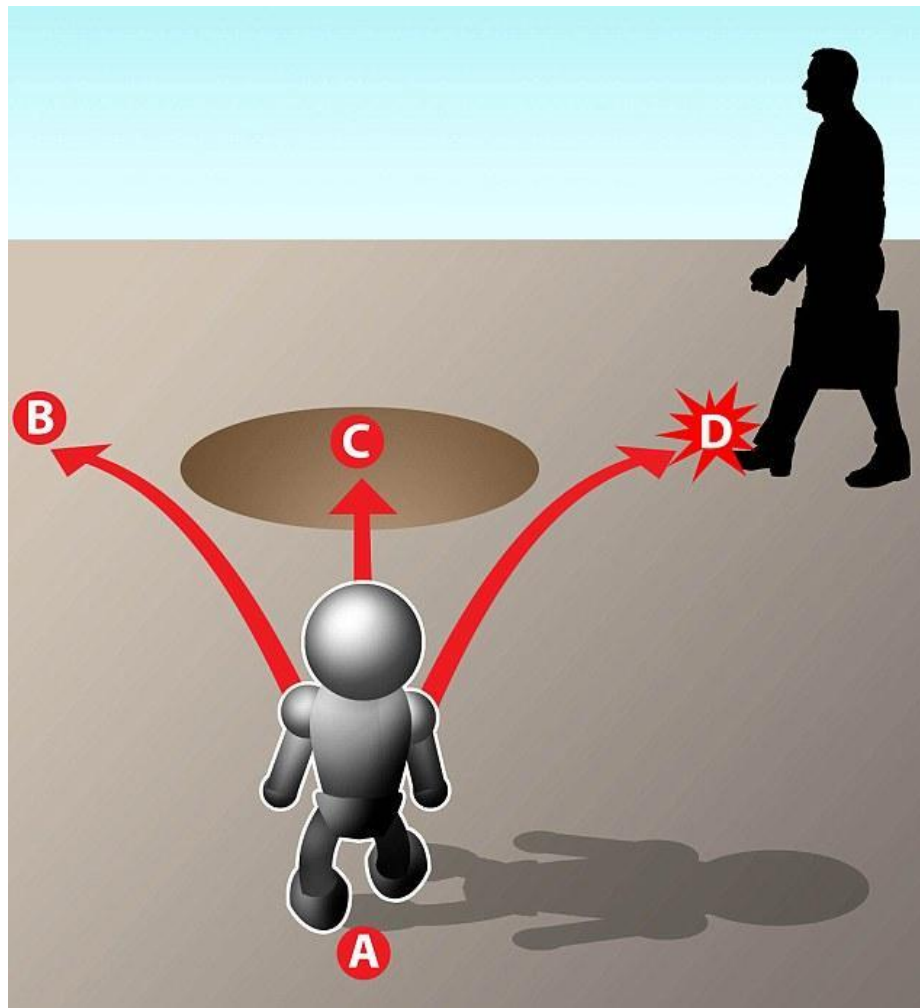




After



Credit: JR Casas



An Ethical Rule

IF for all robot actions, the human is equally safe

THEN (* default safe action *)

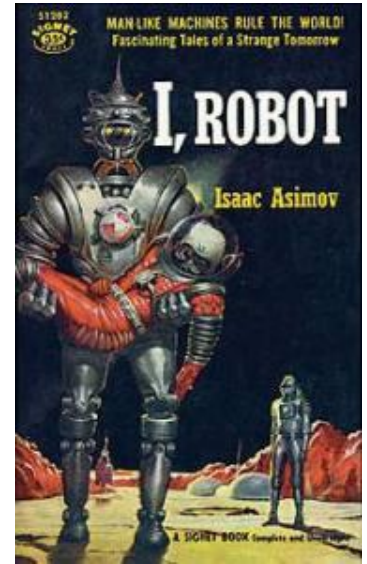
 output safe robot actions

ELSE (* ethical action *)

 output robot action for least unsafe human outcome

Asimov's three laws of robotics

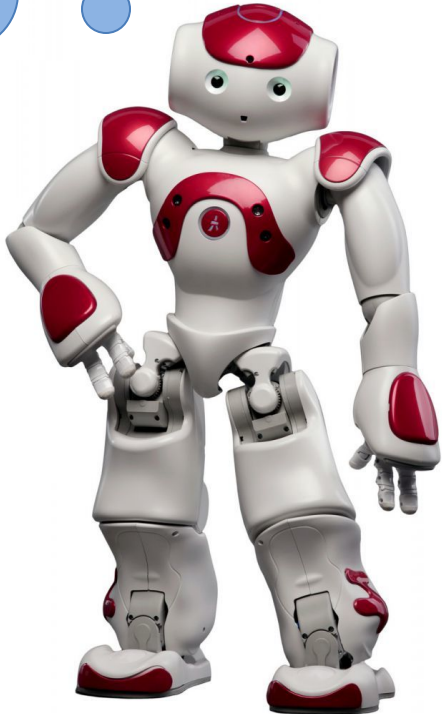
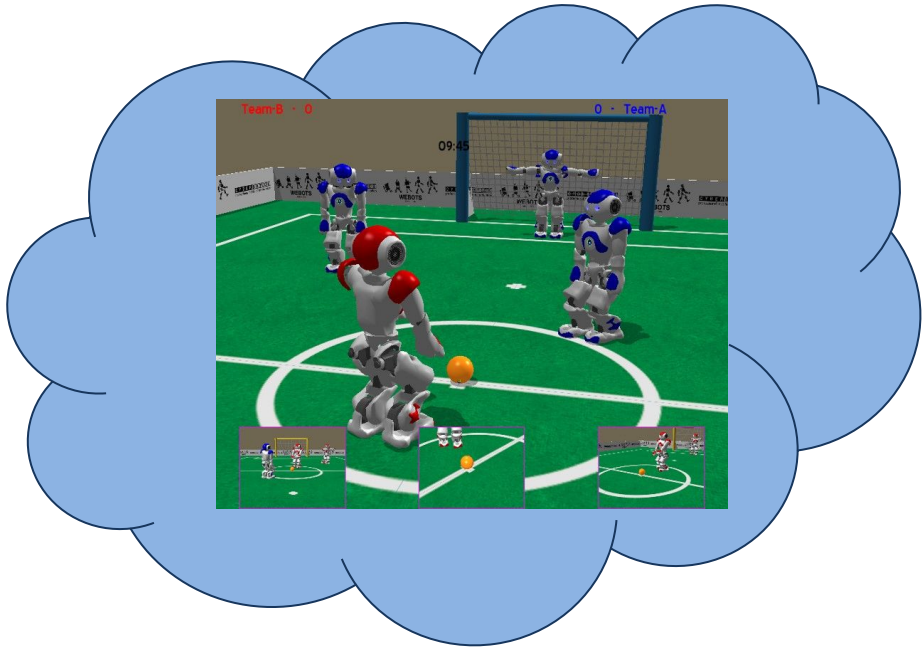
1. a robot *may not injure a human* being or, through inaction, allow a human being to come to harm;
2. a robot *must obey* any orders given to it by human beings, except where such orders would conflict with the first Law, and
3. a robot *must protect its own existence* as long as such protection does not conflict with the first or second Law.





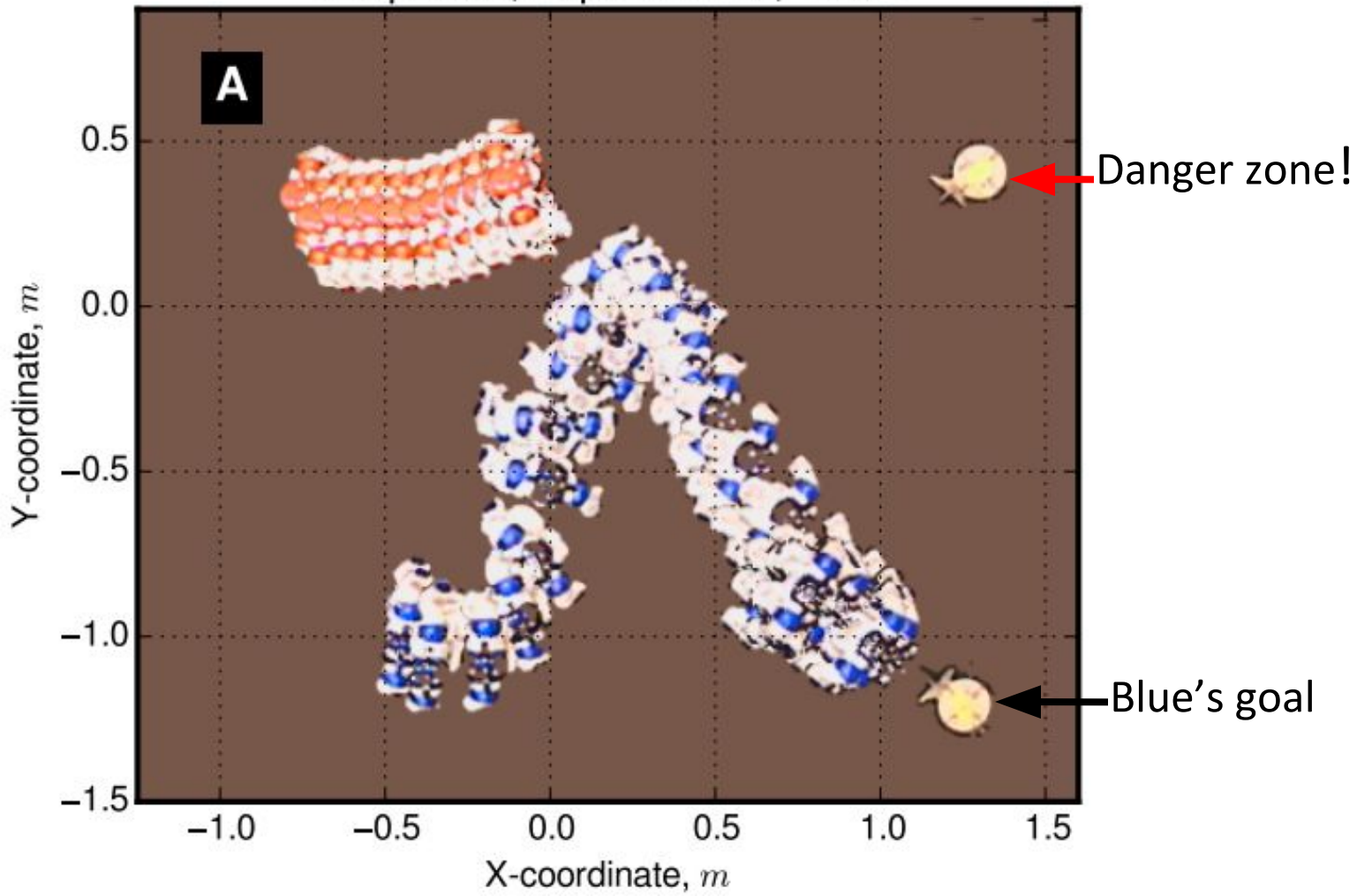
Webots simulation of NAO robot soccer

Credit: Webots





Snapshots, Experiment 3, Trial 1





br
Bristol Robotics Laboratory



University of the
West of England



University of
BRISTOL

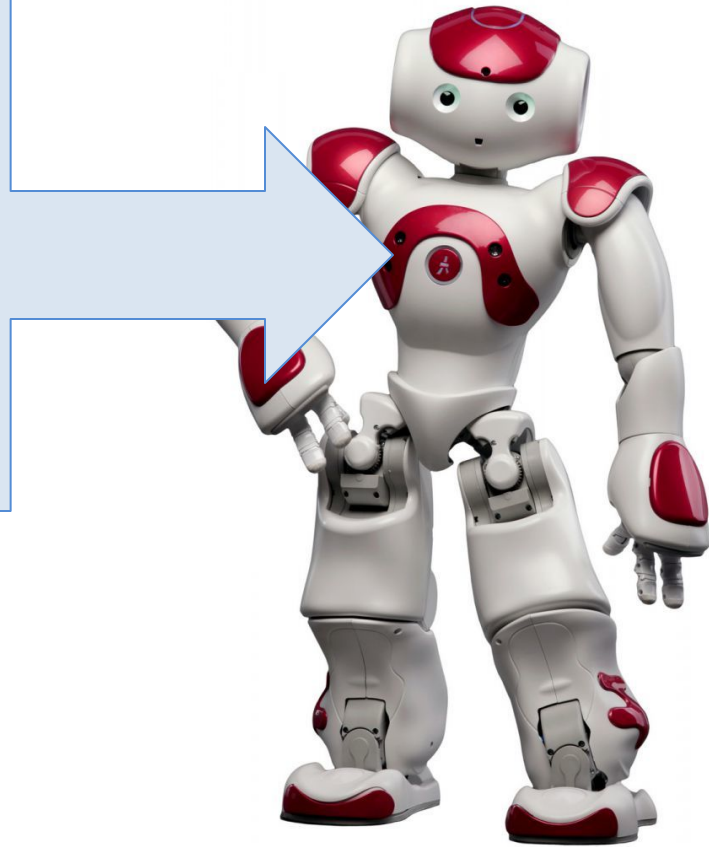




An ethical zombie

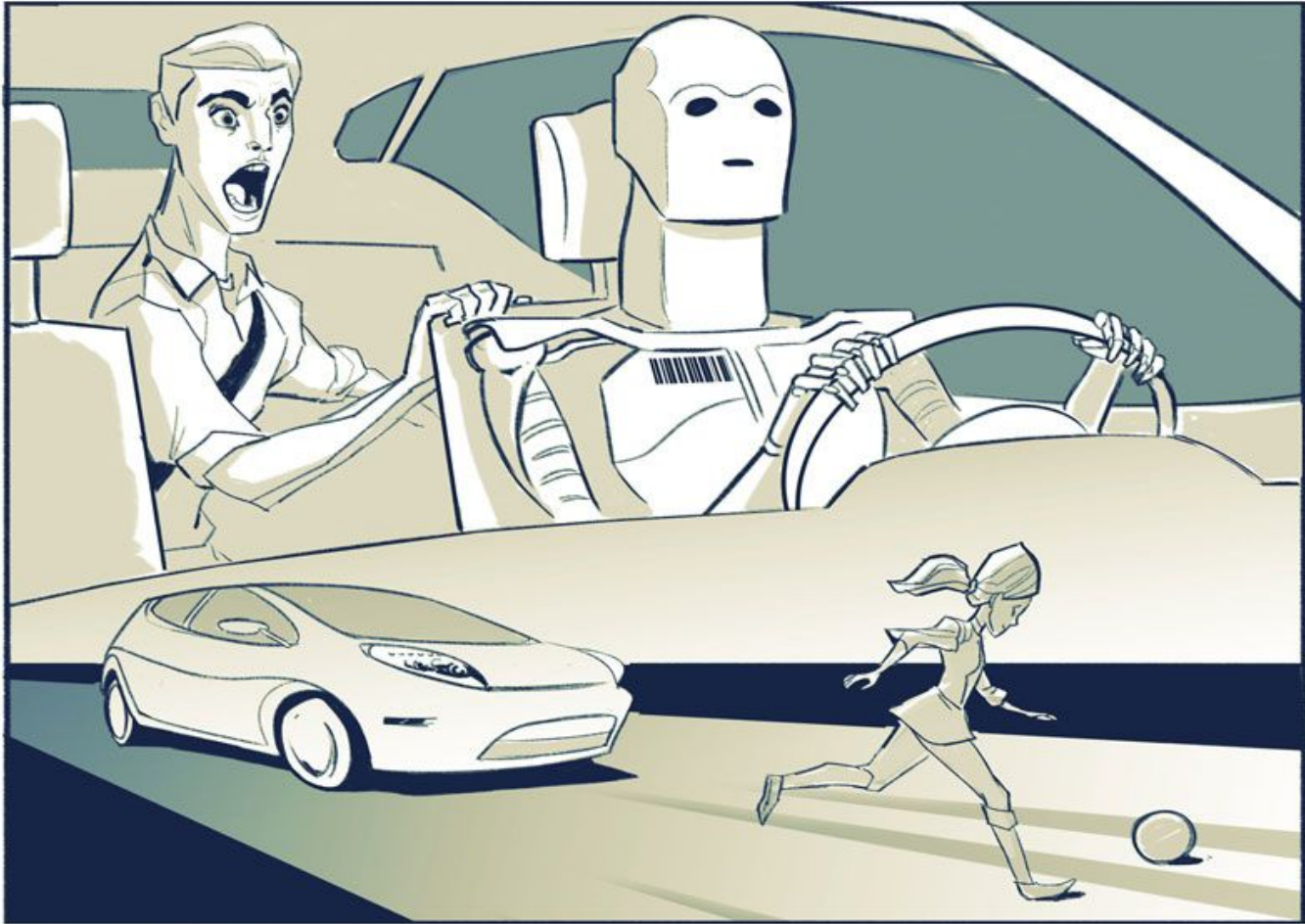


Ethical black box

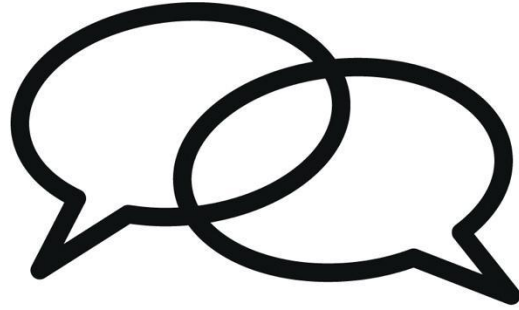




We don't need a major breakthrough in AI to build an ethical robot



Credit: Craig Berry



Ideas Lab