

Self-driving cars: who should AI let die?

Self-driving cars (or Autonomous Vehicles) are in the making. From BMW to Uber, major transport companies are investing in projects aimed at providing with a full-scale autonomous, intelligent car able to drive without anyone directing it.

This would represent a huge revolution for transports: we would not have to worry about parking somewhere downtown or if we had too much wine at dinner, the car will not be affected by it. What will affect the car behaviour however, is the surroundings along which it will ride. Aside from the interaction with other vehicles, the most impelling question is: how would the car react to a situation of foreseeable accident?

A very successful website put online by IMT (<http://moralmachine.mit.edu/>) has been trying to create a number of statistical data showing how we would respond to relatively similar moral scenarios. Building on a famous philosophical thought experiment referred to as the "Trolley Problem" (we will explain more in depth this in the "Neuroscience and the Law" Chapter) the website puts the visitor in a position to choose what in her/his opinion would be the right behaviour (or else, what s/he would do in such a situation) for the car to have. Would you rather run over two old ladies or a healthy young runner? Should one's social status and past behaviour be taken into account when deciding whose life should be spared in such extreme cases?

The website is very interesting because: a) it shows how inconsistent human's choices might be -raising suspects that a case to case moral assessment of the scenario could produce injustice; and b) for the opposing reasons, the experiment shows how a "human eye" could be able to detect important, specific variables that could make the exception to the rule morally acceptable -or even morally required.