Autonomous vehicles (called also self-driving cars or driver-less cars) are one of the emerging technologies that may have a significant impact for society in the upcoming years. Already now the number of companies preparing to manufacture fully autonomous cars is growing, economical and social expectations in that matter are considerable.

The problem is how they large scale implementation can be incorporated into the social life. Although the predictions estimate that the traffic safety will be significantly improved, many people are afraid and prefer human driver's control over vehicles or at least human driver's possibility to take the control over the car. One of the reasons is that people want to be sure that in case of hazardous situation or accident a self-driving car will behave in a proper way. What does it mean "proper way"? This is not an easy to answer question. There are several levels that can be considered here. At the end, however, there is a level of values, especially *moral values*.

In most cases it is possible to avoid damage in property, health and life of passengers and other participants of traffic. It seems credible that a well trained algorithm will perform here better than an average human driver or even a very good driver. However, there are some situations in which some sacrifices are inevitable. For example a self driving car may be forced to decide whether to jeopardise the safety of its passengers or passing by pedestrians. There may be different points of view on such a situation and we do not claim to bring about the only right answer.

Instead, what we believe is crucial for autonomous vehicles' designers is to make clear what hierarchy of values they impose on their vehicles. That will enable the potential owners and users of self-driving cars, other traffic participants, and public in general to accept or reject the wide scale usage of such vehicles. That is also important from the point of view of legal regulations for the area.

The main research hypothesis of this proposal is that formal modelling of self-driving cars and their environment, especially modelling using the tools of logic is a useful step toward clear specification of expectations concerning the behaviour of autonomous vehicles leading to aforementioned social consensus in that matter.

Thus, we propose the language of logic, especially first order logic and its limited variants specific for knowledge representation such as description logic or ontology web language (OWL) as a tool to specify the elements of environment and their properties, potential risk factors, values and preferences.

Moreover, we believe that the existing research results from the area of classical deontic logic, i.e. logic concerning notions of *obligation*, *prohibition* and *permission* combined with some elements of game theory, formal argumentation studies, preference modelling and defeasible reasoning may be successfully applied to the problems of the intended behaviour of self-driving cars.

The subject requires international collaboration since the social attitudes and legal regulations have to be, to some extent, harmonized internationally. Thus we gather a group of researchers from Europe (Poland, Germany, UK) and China to make sure that different points of view are taken into consideration.