ABSTRACT

With the rise of automated driving systems (ADS)-equipped vehicles, the discussion regarding safety ethics of ADS have increased in the recent decade. Some researchers have discussed to define the specification of ADS that present moral issues during collision. Those argued that the specification is based on the least amount of harm is the most ethical consideration while other considered it depends on the societal norms. A standard specification is needed to define a safe and ethical ADS that meet the corresponding level of safety and accountability. In this, philosophy concepts can be used to translate ethical principles into standard specifications.

Recent advances in philosophy for technology have brought a new control concept, known as Meaningful Human Control (MHC). MHC's basic idea is to enable a more morally responsible design of ADS, while also considering the driver and vehicle capabilities. Even in the case of the absence of any specific type of operational control from a human, MHC maintains the relationship conditions between controlling agents and controlled systems that preserve moral responsibility and clear human accountability. Tracing and tracking are two formal requirements that MHC relies on. Tracking is a condition which measures responsiveness and intentions of a system according to a moral reason of human. Tracing, on the other hand, is a condition in which at least one or more human agents appreciate the system and recognize that they are the target of the system's moral implications.

As a promising philosophical concept, several researchers have used MHC to reveal the existence of control gaps that are not obvious from the perspective of normal operational control while others have developed this concept more for ADS application. Furthermore, some have attempted to incorporate the more technical aspects of MHC into ADS. One of the first attempts is to operationalize MHC with a mathematical equation with a case of an ego-vehicle overtaking a cyclist. The most recent one is the definition of four actionable properties for AI systems under the tracing and tracking definition.

However, the existing MHC is still a general concept, though several researchers have started to narrow it down for certain applications and developed mathematical equations, none of them focus on the real implementation of MHC to the existing ADS control system algorithm. The objective of this research is to develop an implementable MHC for ADS through control theory. One of the main ideas is to quantify human reasons and driver capacity & knowledge as a reference for the control system, whereas the control system methods will be determined by finding the method that best similarity the MHC concept.