Meaningful Human Control over Automated Driving Systems

Heikoop, Daniël; Hagenzieker, Marjan; Mecacci, Giulio; Santoni De Sio, Filippo; Calvert, Simeon; van Arem, Bart

Meaningful Human Control over Automated Driving Systems

MHC-ADS

Introduction

With automated vehicles becoming increasingly autonomous, with systems such as adaptive cruise control (ACC) or lane keeping assist (LKA), and traction control or anti-lock braking systems (ABS), the driver is getting ushered more and more away from the classic driving task and towards a new role of system supervisor. Therefore, we need a new way to ensure drivers to remain in control of their vehicles. The notion of *meaningful* human control, derived from the autonomous weapons domain, appears to offer an elegant solution to this progressively urgent issue.

Defining Meaningful Human Control

The definition of Meaningful Human Control (MHC) over automated driving systems (ADS) distinguishes two conditions, namely *tracking* and *tracing*.

Tracking

The ability for a decision-making system (such as ADS) to at all times be responsive to (i.e., 'to track') the human agent's (e.g., driver) relevant reasons to act, ranging from regarding the destination one is to go, to conventional social moral reasoning.

Tracing

The possibility to trace an ADS' behaviour back to some (responsible) human agent (e.g., operator, supervisor, designer, etc.). At least one human agent within the system's design- and use history needs to be able to understand both the capabilities of the system, and their own role as targets of potential moral consequences for the system's behaviour.

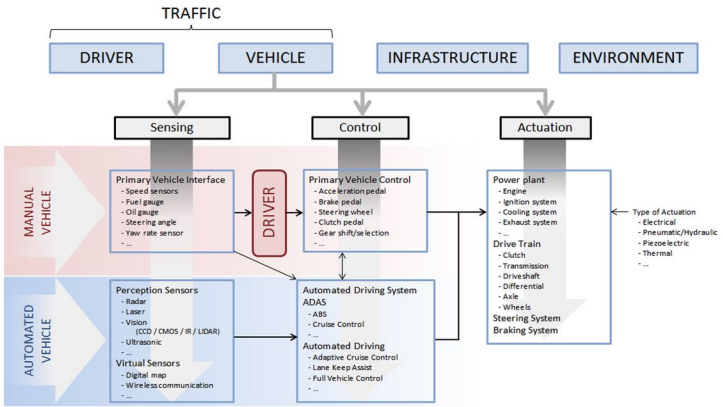
Exploring human control over ADS

Controlling a(n automated) vehicle requires:
- skill- (steering the wheel)
- rule- (following a speed sign)
- knowledge-based behaviour (navigating in a new town)

How many are actually needed, and how are they affected by automating the driving task?

Core components of ADS with MHC

Driver, Vehicle, Infrastructure, and *Environment*: these are the categories wherein core components relevant to ADS have been identified.



Automation / Human	SAE 0 No automation	SAE 1 Driver assistance	SAE 2 Partial automation	SAE 3 Conditional automation	SAE 4 High automation	SAE 5 Full automation
Skill	128	127-114	114	43	40-0?	39-0?
Rule	254	255-250	250	69*-66	51-29?	29-0?
Knowledge	64	64-80	80	33?!	0-?!	0?

Assessing and implementing Meaningful Human Control in ADS

Meaningful Human Control over ADS implies:
- full awareness (or traceability) of the systems' status and actions
- an ADS that knows what you want (i.e., trackability)
- addressing several, if not all, core components involved
- overcoming the unavoidable and undesired shifts in skill-, rule, and knowledge- based behaviour
- developing a smooth transfer of control over various levels of automation

Only if we address the definitions, components and behaviour mentioned here, we will be able to implement a Meaningful Human Control over Automated Driving Systems.

Daniël D. Heikoop
Marjan Hagenzieker
Giulio Mecacci
Filippo Santoni de Sio
Simeon Calvert
Bart van Arem

contact: d.d.heikoop@tudelft.nl
website: www.tudelft.nl/citg/mhc-ads

Meaningful Human Control over automated driving systems

TU Delft