connected automated driving.eu

Breakout session 1:
Large AD pilots

Stakeholder workshop 4: Common Evaluation Methodology
23 November 2020

ARCADE is funded by
the European Union Horizon 2020
Work Programme

Agenda

- Introductions
- Padlet wall (padlet.com/satuinnamaa/7f7goi3xw8j5wszb)
  1) What should be in the guidance given in CEM?
  2) What should not be in the guidance given in CEM?
  3) What can your project provide for CEM?
  4) Which gaps we still have in methodologies that should be developed for CEM?
- Wrap-up

Note

- *There are no right or wrong answers, just different views.*
- *The content of CEM will be decided by the team that sets it up. This material will be provided to them to support their work.*

connected automated driving.eu

What should be in the guidance given in CEM?

- **Common minimum dataset (not mandatory)**
- **Common glossary (incl. evaluation goals)**
- **Methodology structure and common KPIs**
- **Addressing uncertainties and variabilities**
- List of common methodologies
- Established evaluation results (where does it work and where doesn't it?)
- FESTA methodology structure
- Reporting structure
- Lessons learned regarding baselines
- Cost-benefit analysis

What should not be in the guidance given in CEM?

- Too many details
- Mandating instructions of methods / How evaluations should be conducted (leave room for development)

What can your project provide for CEM?

- **LEVITATE: Scalability and transferability of results**
- **L3Pilot: data format, FESTA for AD pilots, options for baseline, practical guidance for running pilots**
- Lessons learnt (pitfalls, do's and don'ts)
- C-Roads CEM, evaluation studies of C-ITS
- UK national methodologies

Which gaps we still have in methodologies that should be developed for CEM?

- **Upscaling**
- **How to construct future scenarios (and traffic and driving scenarios)**
- **Combining different impacts together from different methods**
- **Largely agreed safety indicators (how to go from KPIs to safety impact)**
- Commonly agreed way of modelling AVs in commercial simulation tools
- Common approach around Europe
- Data sources for describing traffic, accidents etc. on EU level, incl. all dimensions of ODD
- How to include the end user in the evaluation (if test drivers need to be used)

Break-out 2: Small and national projects

Stakeholder workshop 4: Common Evaluation Methodology
23 November 2020

ARCADE is funded by
the European Union Horizon 2020
Work Programme

Conclusions BO2: Small and national projects

My project could contribute to CEM:

- Reference / common scenario database
- Methodological material for evaluation
- Liaison national results
- Connectivity element in scenarios
- Virtual risk assessment

Differences between large and small or national projects

- National/smaller projects more focused, more flexible and agile
- Bigger projects have more slack in the planning

Conclusions BO2: Small and national projects

The Common Evaluation Methodology should include:

- KPIs
- Shared model for coming from measurement to evaluation
- step by step process model
- Best practices
- Shared evaluation needs across countries

And not: technical tests, work under standardisation

The Common Evaluation Methodology should address these gaps:

- Advice on common issues
- Evaluation of safety benefits of AD and other socioeconomical issues
- Access to sensitive data (data sharing) or guidance how to arrange this with industry
- adaptable to different test scales
- Agile methods, dealing with project issues

connected automated driving.eu

connected automated driving.eu

Wrap-up
Breakout session 3:
Simulations and data

ARCADE is funded by
the European Union Horizon 2020
Work Programme under the grant agreement No. 824251

Gap 1: Accurate integration of safety strategies in existing traffic simulation software

- Traditionally, the behaviour of human driver agents is only slightly modified in simulations, to include changes in driving behaviour due to using a new system. However, automated driving can differ more: computers have different strengths and weaknesses when compared to humans
- Safety strategies of automated driving manufacturers are multilayer calculations, where the vehicle monitors safe longitudinal and lateral gaps to other road users, optimises path etc.
- The strategies are based on interpretation of traffic rules, but the traffic rules themselves aren't very mathematical: they require, e.g., to approach children carefully, but what is that in meters/speed?
- Mobileye's RSS (Responsibility Sensitive Safety) is so far one of the only definitions of mathematically safe driving, but the development is still in the beginning. Most other approaches haven't been published.
- Especially lateral control when passing other road users, or speed at busy intersections, should be modelled to accurately represent the cautious driving style of AVs.

Gap 1: Accurate integration of safety strategies in existing traffic simulation software - DISCUSSION

- Variety of simulation tools, sometimes even within one project
- Initial frameworks to standardize models
- Panel input: functional requirements for AVs, regardless of technology. Simulate such functionality.
- Certain simple logic can be replicated: no overtaking, fixed path
- Data from near-misses or collisions, vehicle reaction in those, helps to model
  - Difficult to get from prototype testing but could come from other sources like technical testing, disengagement reports, user reporting?
- Passenger modelling, bus stops and stations

Gap 2: Edge cases

- What are they (definitions)?
  - Deep puddle? Snowmobile? A drunken pedestrian? A parachutist landing accidentally?
- How to find them?
  - Disengagements? Tesla's beta has a report button
- How to deal with them?
  - Constantly expecting something unlikely to happen, even a cyclist falling or a child jumping from between parked cars, can lead to very cautious driving – or even to requirements about having an own free lane. To what degree should an automated vehicle include unlikely accidents in its safety margins?
  - Deep understanding of uncommon situations (flooded road, major accidents) is outside the usual scope of navigation systems, but not necessarily impossible with extra scenario-related detection and logic. Merely just stopping can be effective, but the vehicle shouldn’t block the road.
  - Monitoring e.g. road conditions is technically difficult

Gap 2: Edge cases - DISCUSSION

- Exact definition difficult: start, stop, frequency. Difficult unknowns/unseen that could cause danger?
- Link to scenario testing, fail/pass criteria
- Do we need to test all of them?
- Lack of data is a clear difficulty. Valuable for training e.g. AI
- How much should the automated vehicle prepare for something very unlikely
  - How well a human driver can deal with a certain edge case, compare AV performance to that?
  - At least as good as human drivers
- Cases are ODD-related

Gap 3: Simulate automated mobility in the context of urban planning

- Automated driving and new mobility services need to be studied in a wider context than in small pilot tests.
- Urban planning simulations could combine city design goals with automation enablers and bring new insight about different outcomes and help to find out whether and how automation can support achievement of the goals.
- In addition, indirect impacts (reasonably foreseeable impacts that occur later in time or farther away than direct impacts) and long-term effects remain partly hidden in small-scale tests. New types of simulations could be used scale up initial findings and test assumptions and to allow testing how different policy actions affect the outcome.
- Related simulations, for example: dedicated lanes, different parking options, new transport modes, new services, integration with the old

Gap 3: Simulate automated mobility in the context of urban planning - DISCUSSION

- Complex simulation environment, when compared to e.g. highway
- Infrastructure data: cameras, loop sensors
- Comes back also to physical and digital infrastructure requirements for automated driving
- Requirements for data from infra, e.g. some data from parking houses, map/position...
- Not just for urban planning but also for traffic system operators

Gap 4: Could new projects share a common log data format?

- Having a common data format enables joint development of post-processing tools, such as quality checking, scenario detection and indicator calculation. A large set of tools and calculations is very expensive to develop again and again from scratch.
- A common format makes it easier to share data between organisations
- A common format can capture the needs of several impact assessment areas
- However, a common format may or may not include confidential data such as video (privacy issues) or performance-related detailed data (confidentiality), raw sensor data
- Using a common format requires conversion from vehicle formats to the format – or directly logging in the format. Conversion efforts may be too demanding for small projects, unless the toolset is very easy to use. Small projects usually collect their data in various formats, e.g. csv.
- L3Pilot has published an open format for the evaluation community. Waymo shares data in their format, but not yet for evaluation purposes.

Gap 4: Could new projects share a common log data format - DISCUSSION

- Common format has large benefits, but how easy to take into use in new projects
- Modularity of the format could be a key, which parts would one use in a project
- Minimum information content enables conversions
- Should be easily adaptable – balance that with tools? Tools for separate modules?
- Post-processing comes with the format or no, it doesn't in OSI[1] or TR68[2]?
  - ARCADE could link these to knowledge-base, even overview of different formats?

[1]: https://github.com/OpenSimulationInterface/open-simulation-interface

[2]: https://www.singaporestandardseshop.sg/Product/SSPdtDetail/1cd3d49e-e896-4f33-b1d0-dda96f55bf74

Gap 5: Video sharing is difficult between project partners and GDPR doesn’t help

- GDPR causes concerns between project partners on what data could be collected and shared.
- Most initial concerns are not usually warranted, as GDPR is not to hinder research. It has several research exemptions.
- Still, the legal situation regarding sharing video even between research consortium partners seems to be a moving target with various views.
- Research projects would welcome EU-level practices and legal templates for sharing e.g. the front view video from a car within a consortium, to speed up work and to avoid lengthy legal preparations.
- The best practice, currently, is to add video data sharing clauses to the consortium agreement, to reduce the need for bilateral agreements? Video sharing also needs to be agreed with test users in their consent forms.

Gap 5: Video sharing is difficult between project partners and GDPR doesn’t help - DISCUSSION

- GDPR doesn't tell to blur but defines clear (at least, after some work) requirements
  - Blurring and such processing can be costly
- ARCADE considering creating a legal summary
- Tools are improving: detection of eye movement, semantic segmentation, … hiding identity

connected automated driving.eu

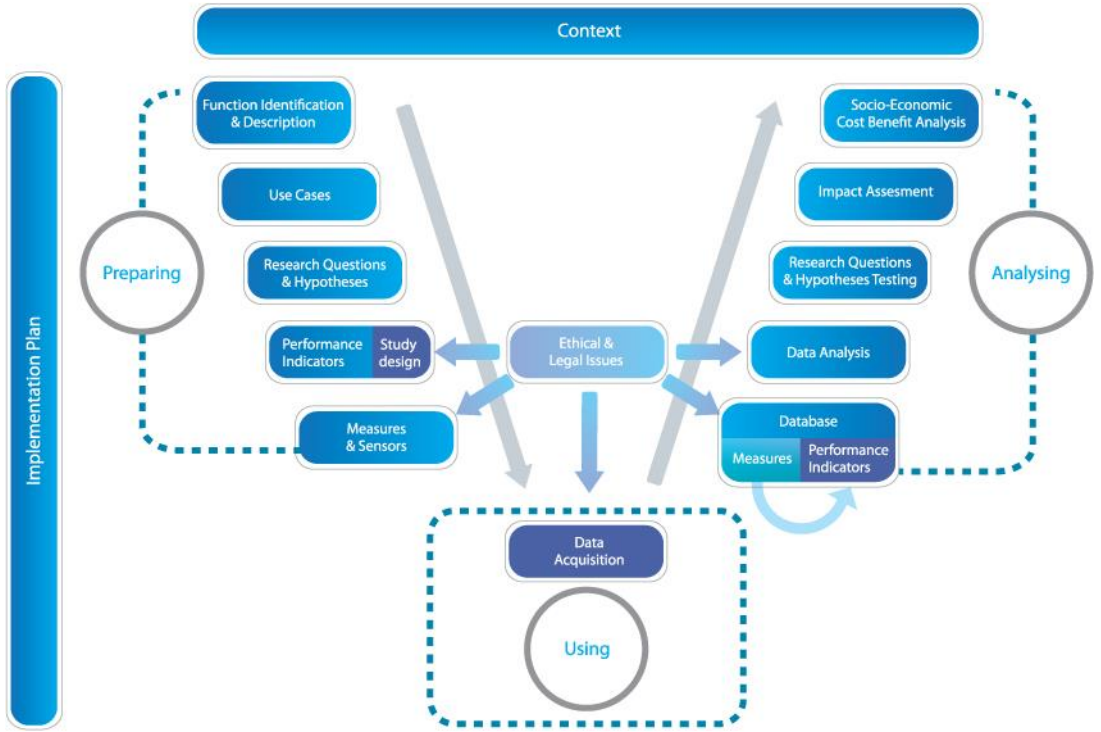
Wrap-up Breakout session 4: FESTA Methodology overview

Stakeholder workshop 4: Common Evaluation Methodology 23 November 2020

ARCADE is funded by
the European Union Horizon 2020
Work Programme under the grant agreement No. 824251

Wrap-up

- Introduction KB
- Overview FESTA
  - Preparing
  - Using
  - Analysing
- Related work:
  - Data Sharing Framework
  - Trilateral Impact Assessment Framework



PREPARING

- Research questions:
  - New research questions arise
  - How to prioritise them?
  - It may be difficult to formulate precise hypotheses
  - Baseline question: What do we compare the outcomes with?
- Restricted scenarios or space mission approach?
- Use cases cannot be only defined by what is desirable to evaluate, but also depends on what is allowed to be tested on the real road
- Not all functions with the same name do the same thing, e.g. highway pilot
- Maybe the study design will need to be more flexible, allowing for exploration, and iteration and revision during the project
- New methods are needed for measuring user acceptance, not only general questionnaires, but capturing user experience, problem: testing is often done with safety drivers

DATA ACQUISITION AND ANALYSIS

- Different sensor types provide different information and data quality
- Automation for data analysis process is necessary to allow analysis of relevant collected data
- Data analysis on distributed data is challenging
- Synchronization among all local data bases should be ensured through automatic procedures
- Novel evaluation methods: visioning, scenario development, data mining, machine learning, automated video analysis, data anonymization, automated scenario detection, …
- How to evaluate AI processes and decisions?
- The main challenge in data acquisition, data processing and data analysis is the establishment of automated processes
- New role of simulations

What is it that we want to learn from FOTs and pilots?

- Impact of road automation on society, users, industry, ….
- The effects on mobility, safety, environment, efficiency, ….
- If possible, also on wider impact areas: land use, health, energy, ….
- The behaviour of vehicle (systems), users, infrastructure, services, ….
- The opportunities and challenges for the future

Why a common evaluation framework:

A common methodology is important for being able to study impact of automation, combining all knowledge gathered from different kinds of FOT and pilot

connected automated driving.eu

Wrap-up Breakout session 5: Framework for efficient operation of automated fleets: the early days

Stakeholder workshop 4:
Common Evaluation Methodology 23 November 2020

ARCADE is funded by
the European Union Horizon 2020
Work Programme under the grant agreement No. 824251

Conclusions BO5 : Framework for efficient operation of automated fleets: the early days.

- What are the concrete challenges and methodological gaps identified by large AD pilots ?
  - Variety of legal schemes, technical challenges, users needs, services provided
- How do we/pilot organisers solve the issues ?
  - Need to learn from each project and share results => link to K B
  - Bring trust into the landscape through a trusted entity
- What are the frameworks we need in priority ?
  - Common ODD, use case descriptions, vehicle denomination
  - Condition for simplified framework : trusted entity
- How can common and harmonised frameworks support in solving the issues / gaps ?
  - Provide effective and efficient testing environments
  - Common pathway for policy makers, be specific on goals and how to measure impacts, effective frameworks for sustainable business, consistency and interoperability
  - Share the learnings for the next phases, we are able to assess maturity, identify the future norms