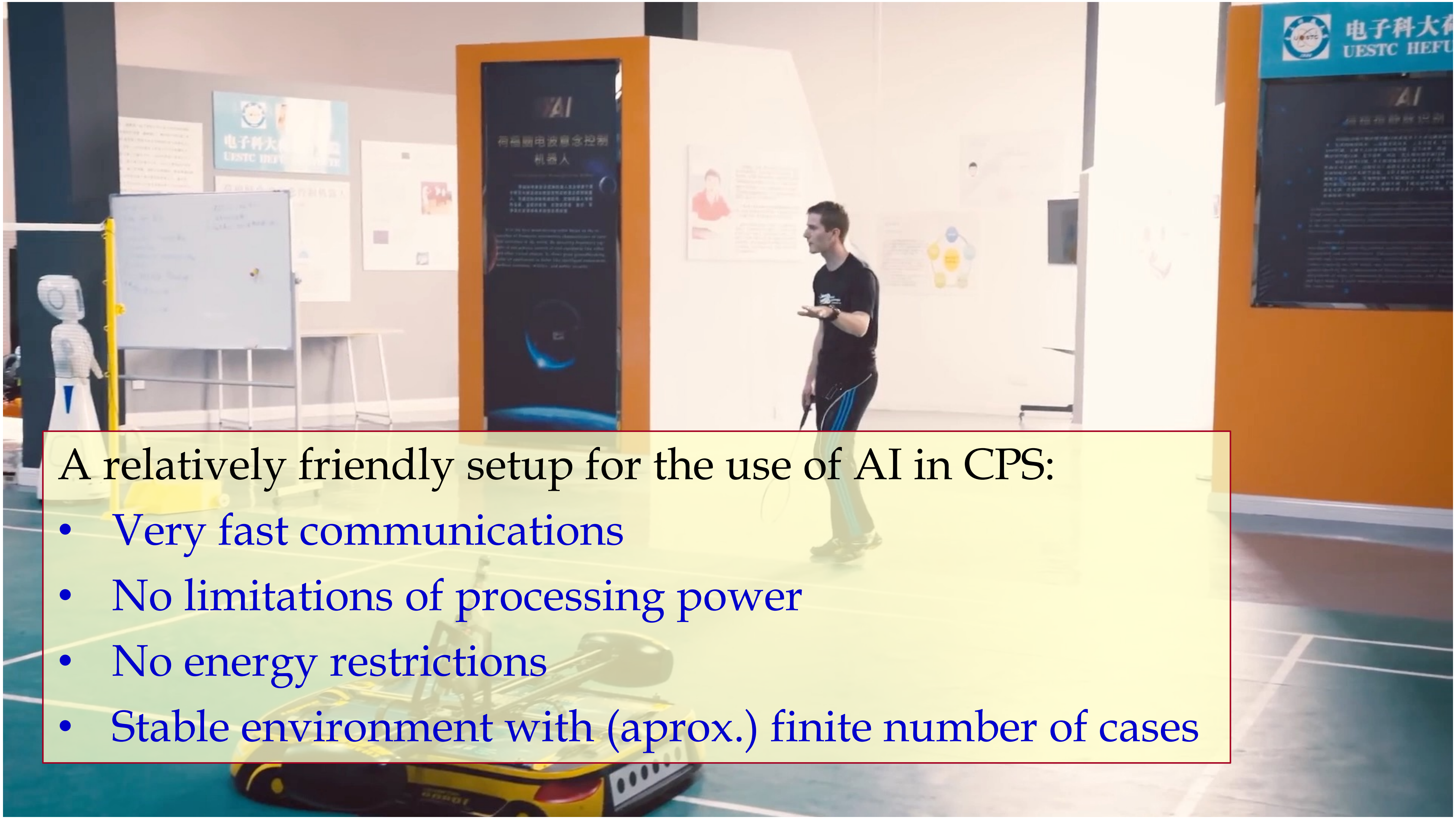
Safety and security in AI enabled critical applications: who is going to solve the problems?

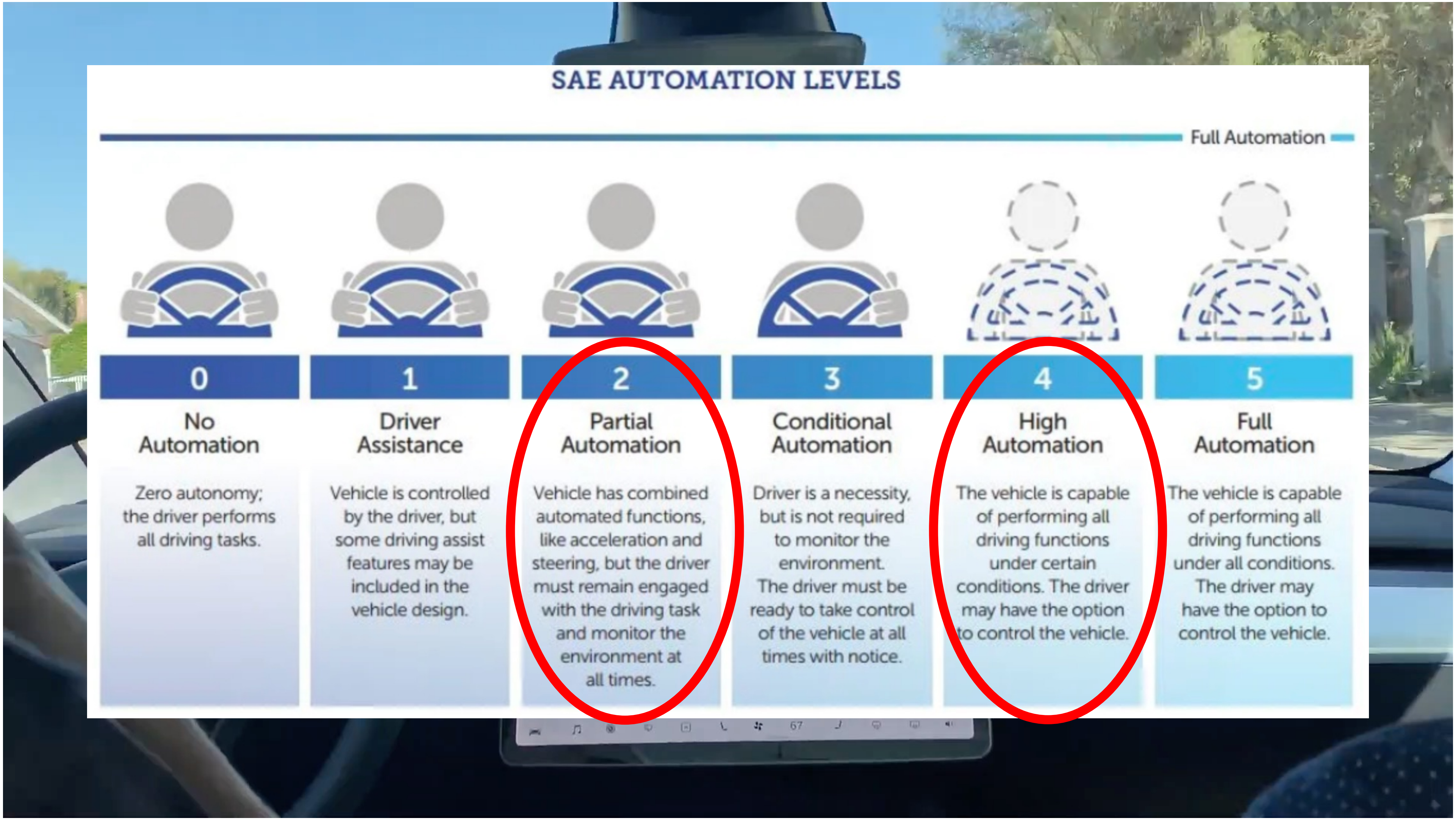
Henrique Madeira
University of Coimbra, Portugal

80th Meeting of the IFIP 10.4 Working Group on
Dependable Computing and Fault Tolerance
Virtual - 25 June 2021 – 27 June 2021



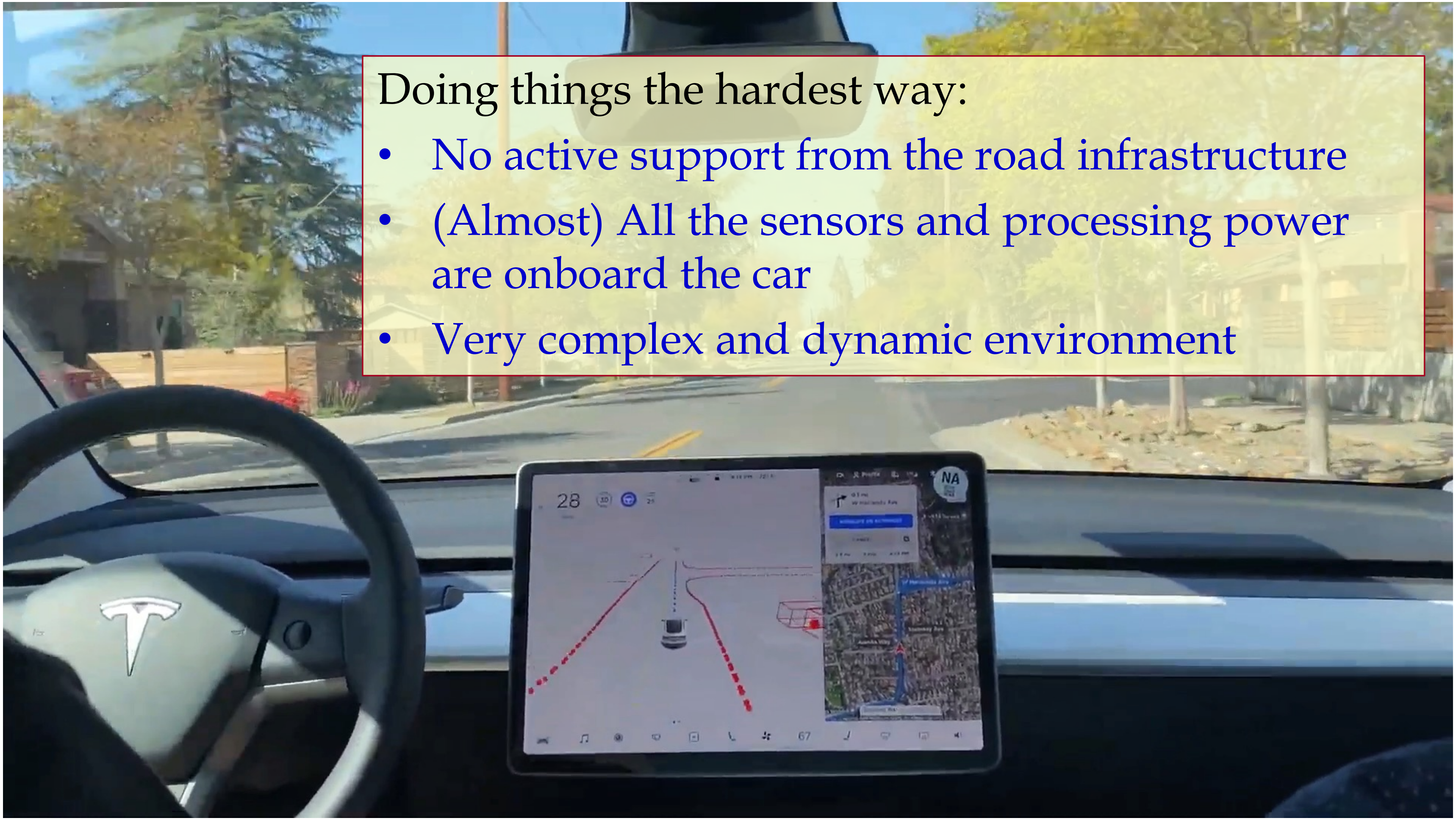
A relatively friendly setup for the use of AI in CPS:

- Very fast communications
- No limitations of processing power
- No energy restrictions
- Stable environment with (aprox.) finite number of cases


SAE AUTOMATION LEVELS
Full Automation
0
No Automation
Zero autonomy; the driver performs all driving tasks.
1
Driver Assistance
Vehicle is controlled by the driver, but some driving assist features may be included in the vehicle design.
2
Partial Automation
Vehicle has combined automated functions, like acceleration and steering, but the driver must remain engaged with the driving task and monitor the environment at all times.
3
Conditional Automation
Driver is a necessity, but is not required to monitor the environment. The driver must be ready to take control of the vehicle at all times with notice.
4
High Automation
The vehicle is capable of performing all driving functions under certain conditions. The driver may have the option to control the vehicle.
5
Full Automation
The vehicle is capable of performing all driving functions under all conditions. The driver may have the option to control the vehicle.

Doing things the hardest way:

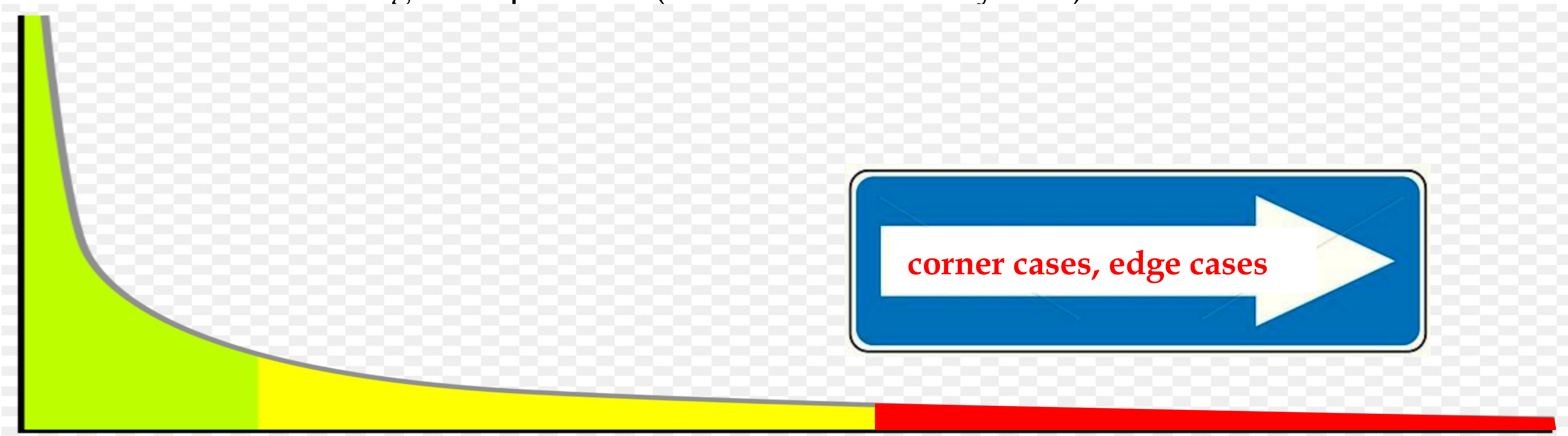
- No active support from the road infrastructure
- (Almost) All the sensors and processing power are onboard the car
- Very complex and dynamic environment



Difficulties in AI enabled critical applications

AI/ML used in safety-critical functions:

- Lack of clear functional specifications
- Non-deterministic and probabilistic outputs
- Limitations of the training data
- Non-explainable ML (i.e., black box)
- Exhaustive testing is impossible (as usual in ordinary SW) but in addition to that ML

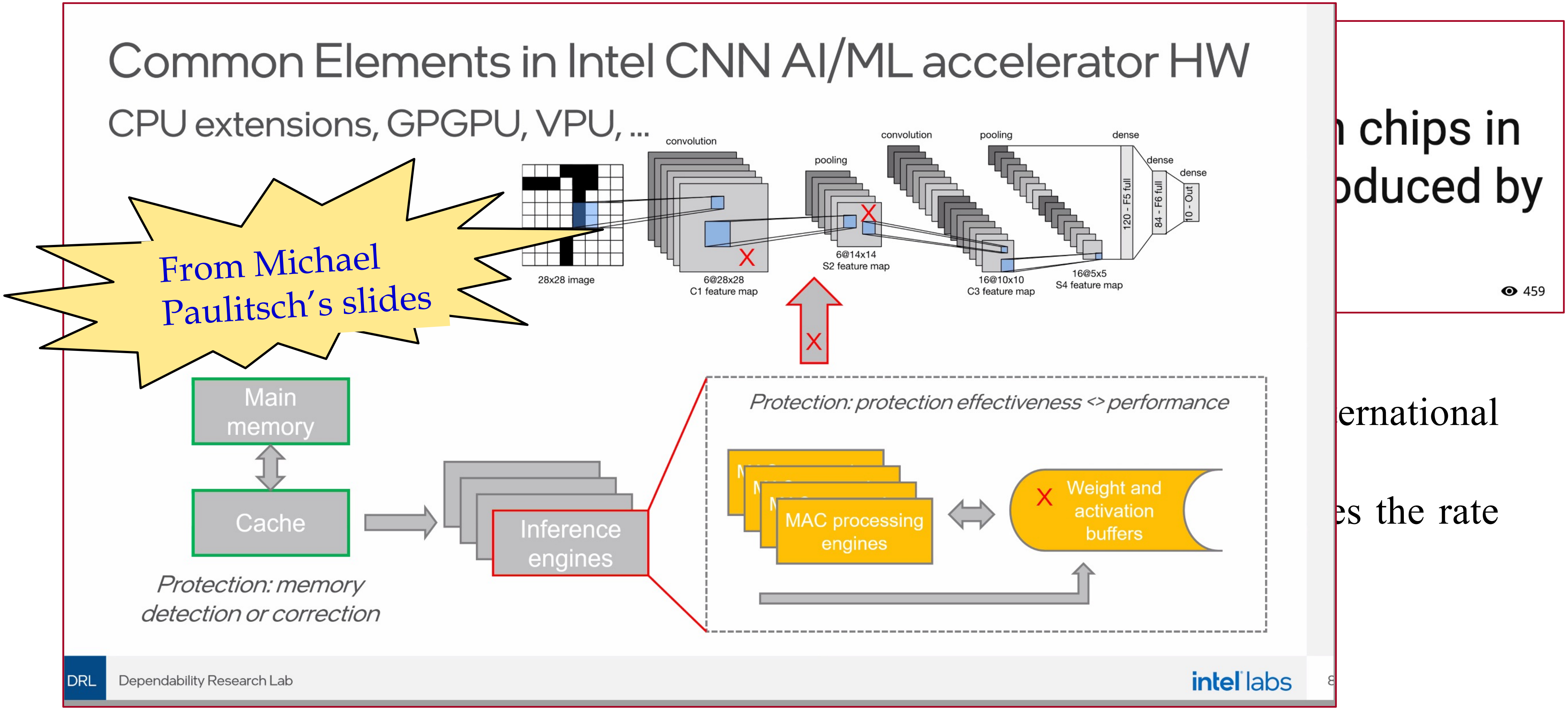


Henrique Madeira, DEI-FCTUC, 2021

Additional (classic) difficulties

- **Software faults**
  - Defect densities remain nearly the same (i.e., high) for decades
  - Many CPS have now millions of lines of code. A modern car, for example, has > 100 millions lines of code.
- **Hardware faults**
  - "Silicon defects are getting worse", Michael Paulitsch (Intel). The International Technology Roadmap for Semiconductors (ITRS) says the same.
  - AI in safely critical applications needs massive hardware ➔ increases the rate of hardware faults

Additional (classic) difficulties



Henrique Madeira, DEI-FCTUC, 2021

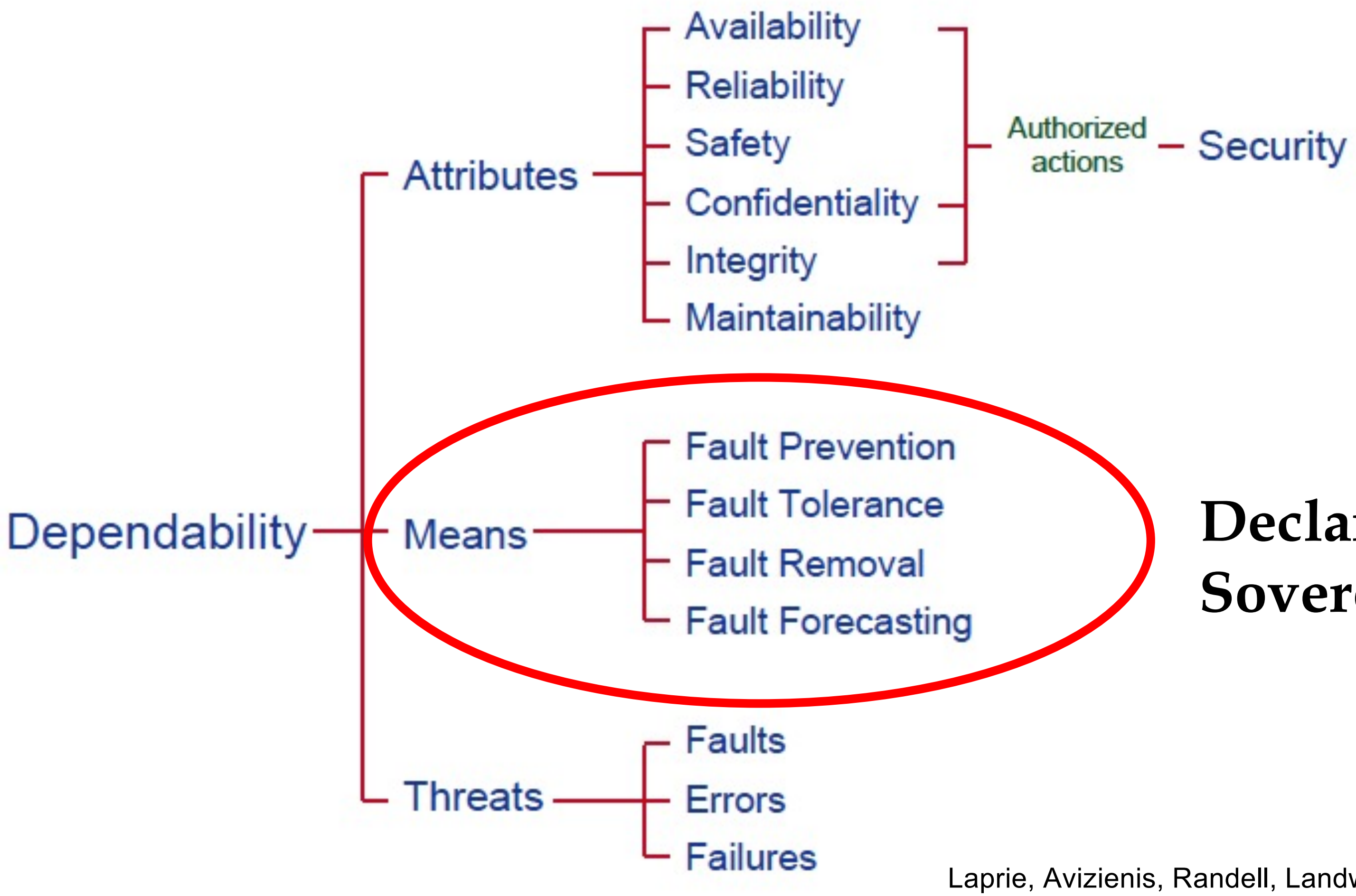
Full self-driving: is it a classic problem?

- **Building dependable systems using components that are not perfect looks like a classic problem**:
  - The output accuracy of AI components (in the absence of faults) is probabilistic (specially for black-box AI)
  - All components are subject to software faults
  - HW faults must be considered
  - AI used in safety-critical applications is an interdisciplinary problem, no matter the application area (automotive, medical devices, industry 4.0, avionics, etc.).

Who is going to solve the problems?

(What is expected from our research community?)

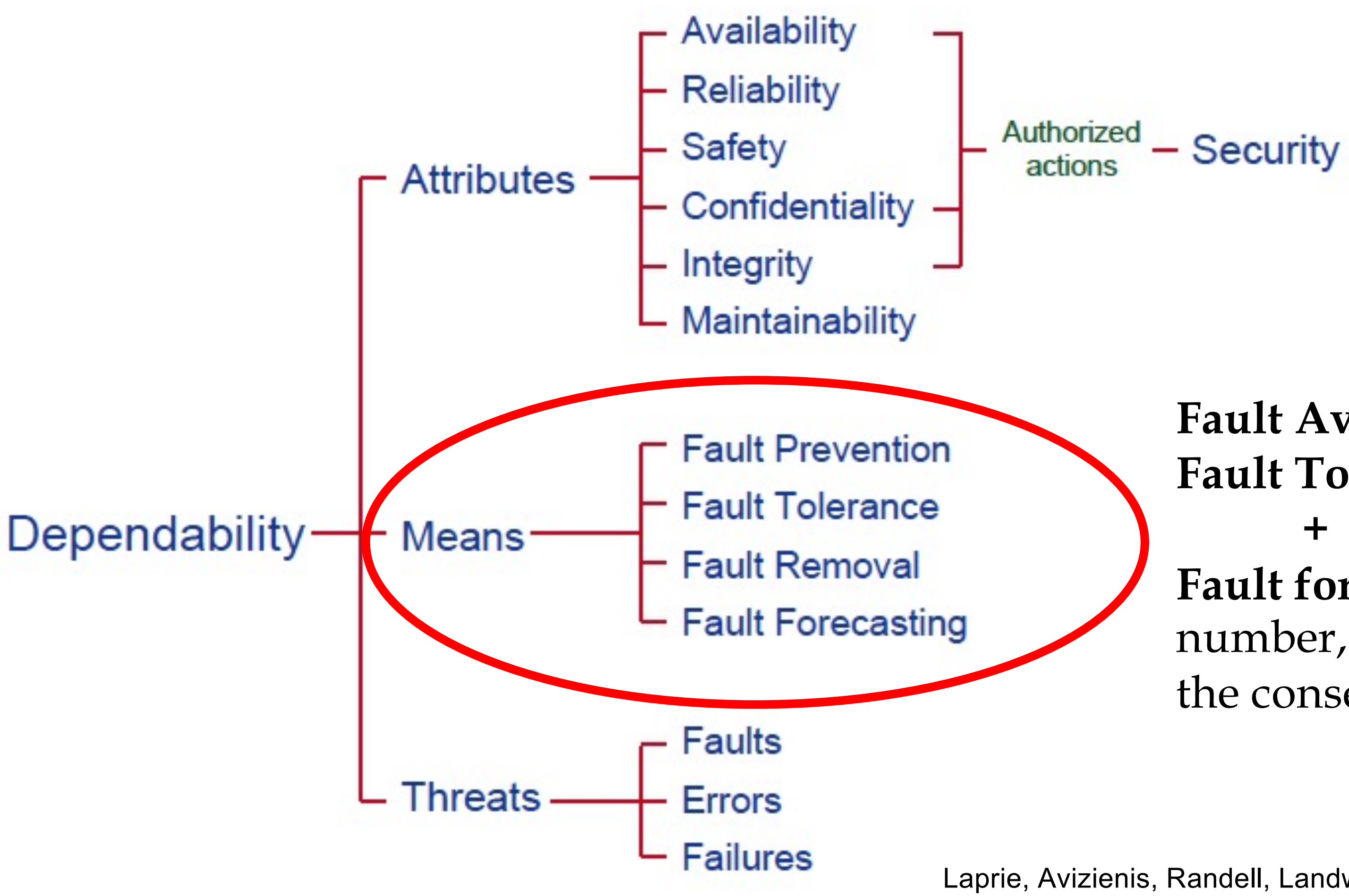
Cartography of (our) Dependability World



Declaration of Sovereignty ☺

Laprie, Avizienis, Randell, Landwehr, Simoncini,…

Cartography of (our) Dependability World



Fault Avoidance
Fault Tolerance
+
Fault forecast (estimate the present number, the future incidence, and the consequences of faults)

Laprie, Avizienis, Randell, Landwehr, Simoncini,…

Who is going to solve the problems?

Problems:

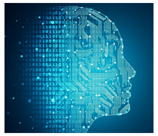
- **How to assure safety and security** in AI enabled safety-critical applications?
- **How to demonstrate** that one can trust on AI enabled safety-critical applications?

Can we do that for self-driving cars?

- Millions of vehicles
- Billions of driving hours
- Huge pressure to cut cost
- Very high criticality

Who is going to solve the problems?

Artificial intelligence



The solution is more and better AI

Problems:

- **How to assure safety and security** in AI enabled safety-critical applications?
- **How to demonstrate** that one can trust on AI enabled safety-critical applications?

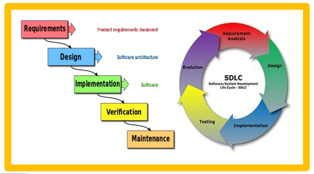
More

- Robust AI models
- Non-symbolic AI
  - Larger training data sets and bigger and more complex neural networks
  - Interpolation vs extrapolation
- Explainable AI
- Ensembles
- …

Who is going to solve the problems?

- A software maturity/testing problem…
- A better process: if you follow these guidelines the result will be good.

Software Engineering



Standards

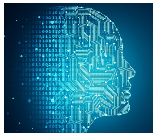
- ISO 26262 - hardware and software failures
- ISO/PAS 21448 – SOTIF
  - Unsafeness due to technology limitations
  - Foreseeable unsafe actions from the users
- UL 4600: Standard for Safety for the Evaluation of Autonomous Products
  - Fully autonomous vehicle (no human driver/supervisor)
  - Safety case oriented standard
  - ”Methodical way to show use of best practices
    - Did you do enough?
    - Extensive lists of #DidYouThinkofThat?”
  - Metrics section

Who is going to solve the problems?

Problems:

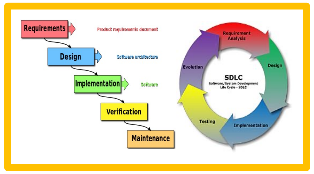
- **How to assure safety and security** in AI enabled safety-critical applications?
- **How to demonstrate** that one can trust on AI enabled safety-critical applications?

Artificial intelligence



The solution is more and better AI

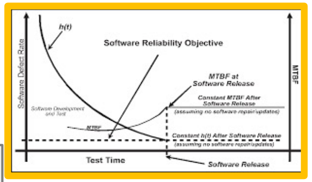
Software Engineering



The solution is in the process

Standards

Software Reliability



The solution is in models and tools

- The goal is to be able to measure software reliability
- Software reliability growth models… and AI

Who is going to solve the problems?

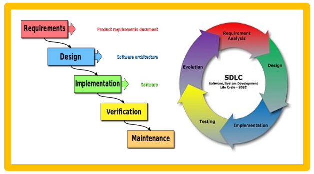
Problems:

- **How to assure safety and security** in AI enabled safety-critical applications?
- **How to demonstrate** that one can trust on AI enabled safety-critical applications?

Artificial intelligence

The solution is more and better AI

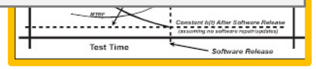
Software Engineering



The solution is in the process

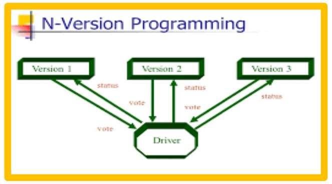
Standards

- Build dependable systems with unreliable components
- Can we use old recipes?



The solution is in the architecture

Dependability



Henrique Madeira, DEI-FCTUC, 2021

Architecture as (part of) the solution

From Yair Amir's slides

Black-Box Monitoring

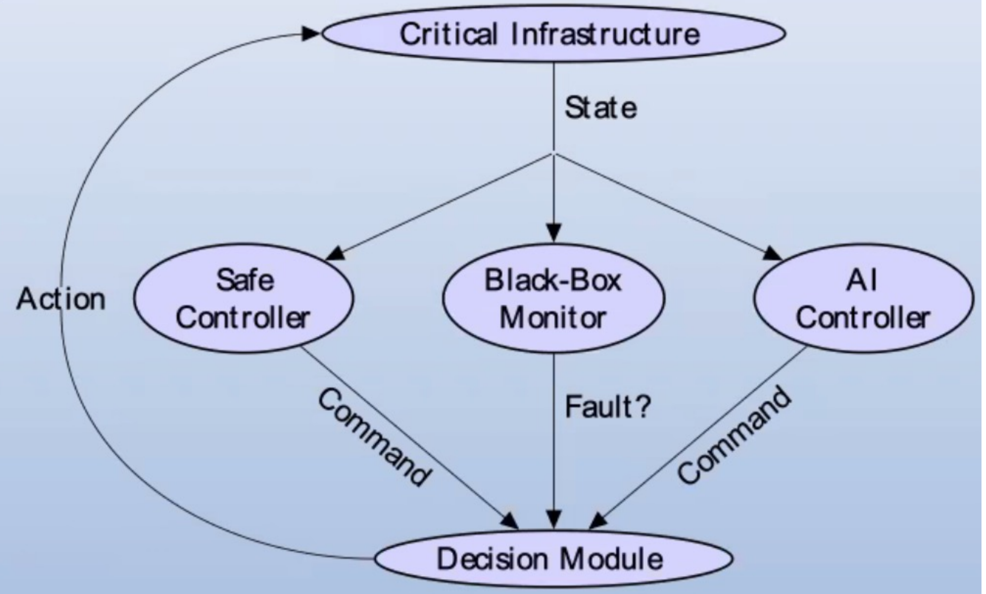
Black-Box monitoring is a standard approach to create dependable systems

The systems work roughly as follows:

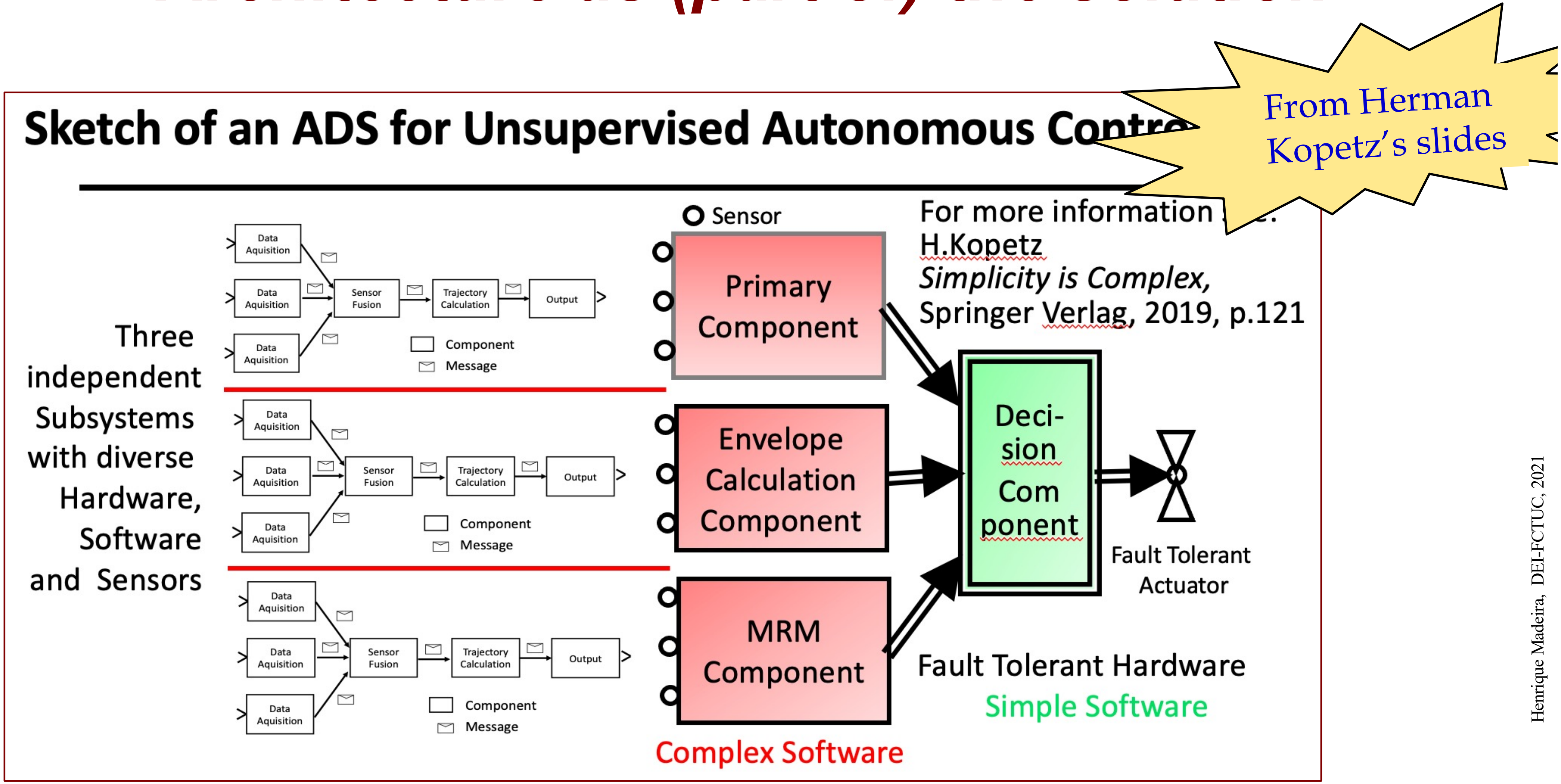
State is collected and passed to a trusted controller, an AI controller and a monitor.

Each controller proposes an action

The decision module uses the output of the monitor to determine which action should be performed



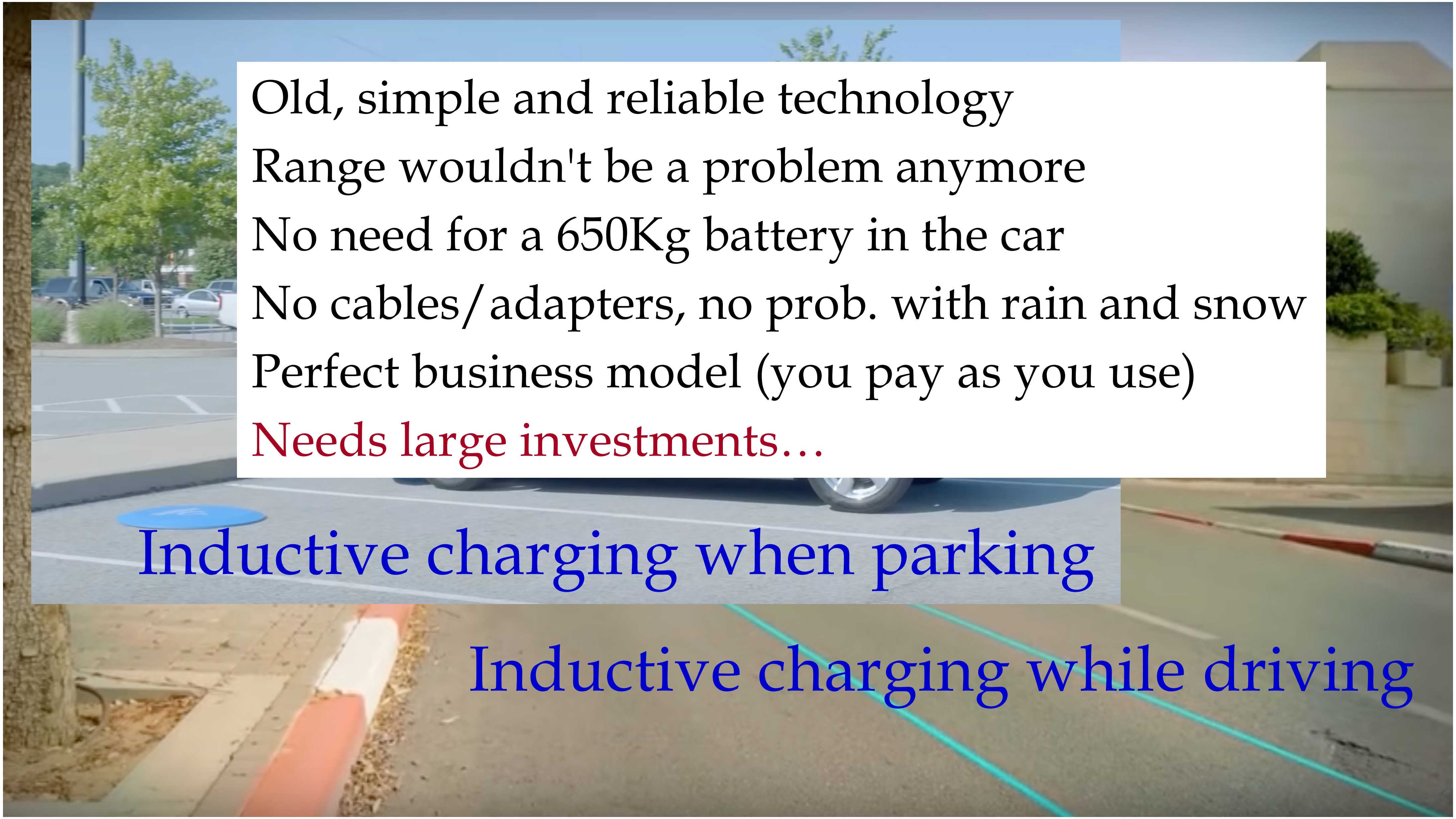
Architecture as (part of) the solution

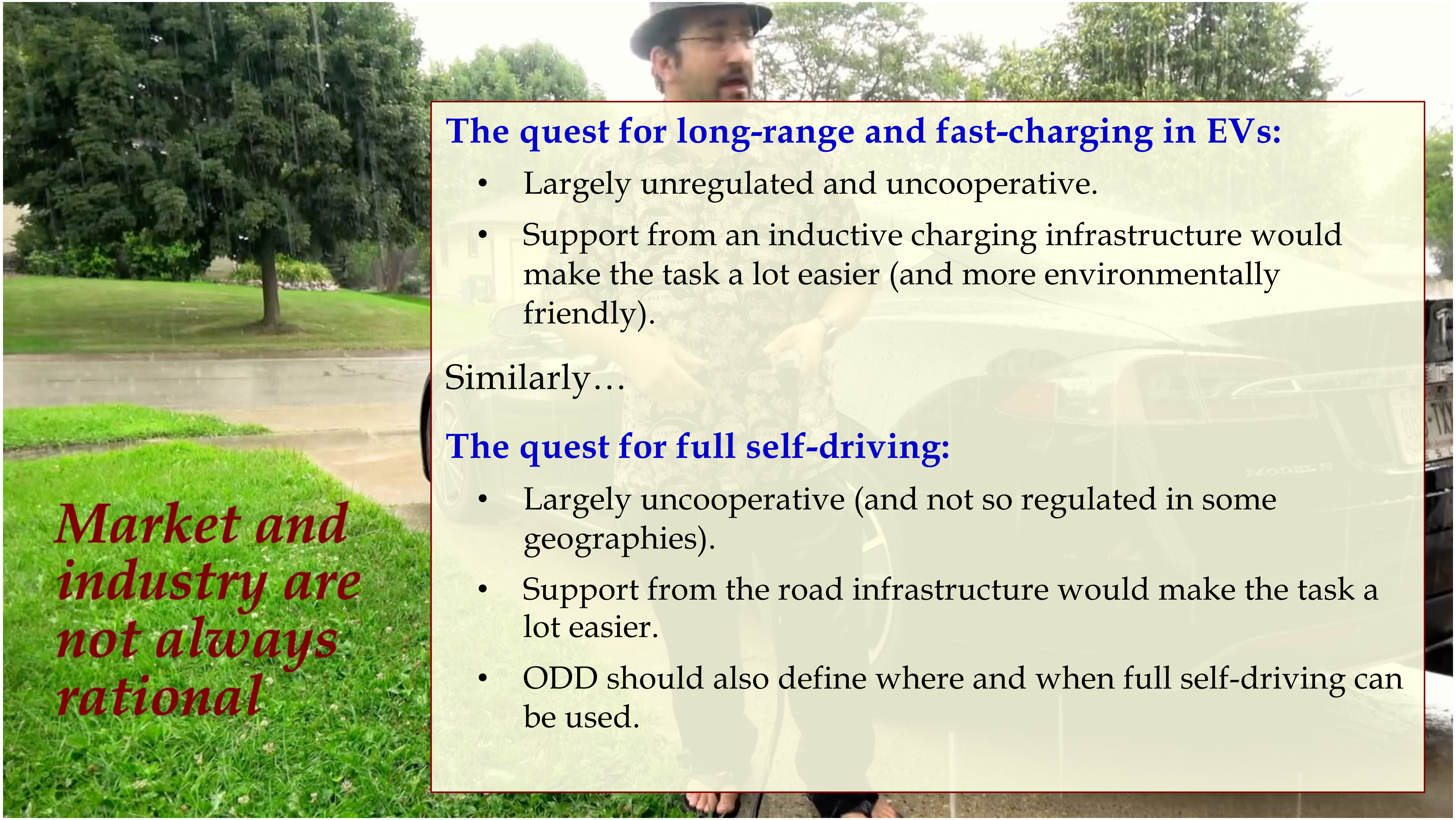


Henrique Madeira, DEI-FCTUC, 2021


Market and industry are not always rational
MODEL S


PHARMACY
Inductive charging when parking
Inductive charging while driving


Old, simple and reliable technology
Range wouldn't be a problem anymore
No need for a 650Kg battery in the car
No cables/adapters, no prob. with rain and snow
Perfect business model (you pay as you use)
Needs large investments…
Inductive charging when parking
Inductive charging while driving



The quest for long-range and fast-charging in EVs:

- Largely unregulated and uncooperative.
- Support from an inductive charging infrastructure would make the task a lot easier (and more environmentally friendly).

Similarly…

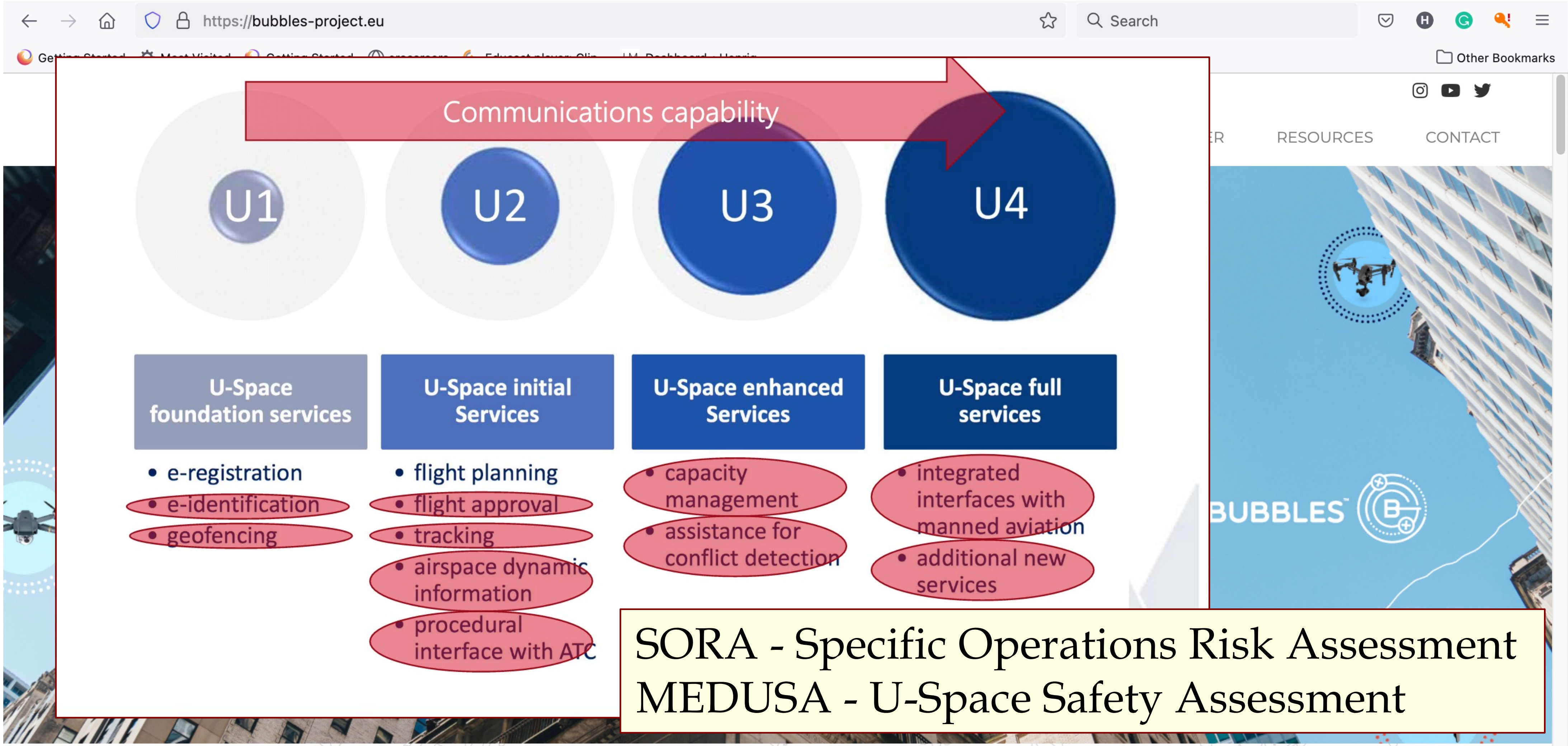
The quest for full self-driving:

- Largely uncooperative (and not so regulated in some geographies).
- Support from the road infrastructure would make the task a lot easier.
- ODD should also define where and when full self-driving can be used.

Market and industry are not always rational


The
Simil
Market and industry are not always rational
NORTH
41
WEST
70S
NORTH
231
EAST
96
Tennessee
SOUTH
231
SOUTH
41
EAST
70S
GIVE WAY TO PEDESTRIANS
ONE WAY
NO ENTRY
STOP
GRAMERCY PARK
SOUTH
ONE WAY

BUBBLES: *Defining the BUilding Basic BLocks for a U-Space SEparation Management Service*



SORA - Specific Operations Risk Assessment
MEDUSA - U-Space Safety Assessment

Full self-driving: is it a classic problem?
(a sort of conclusion)

- **Building dependable systems using components that are not perfect looks like a classic problem**:
  - The output accuracy of AI components (in the absence of faults) is probabilistic (specially for black-box AI)
  - All components are subject to software faults
  - HW faults must be considered
  - AI used in safety-critical applications is an interdisciplinary problem, no matter the application area (automotive, medical devices, industry 4.0, avionics, etc.).

Who is going to solve the problems?

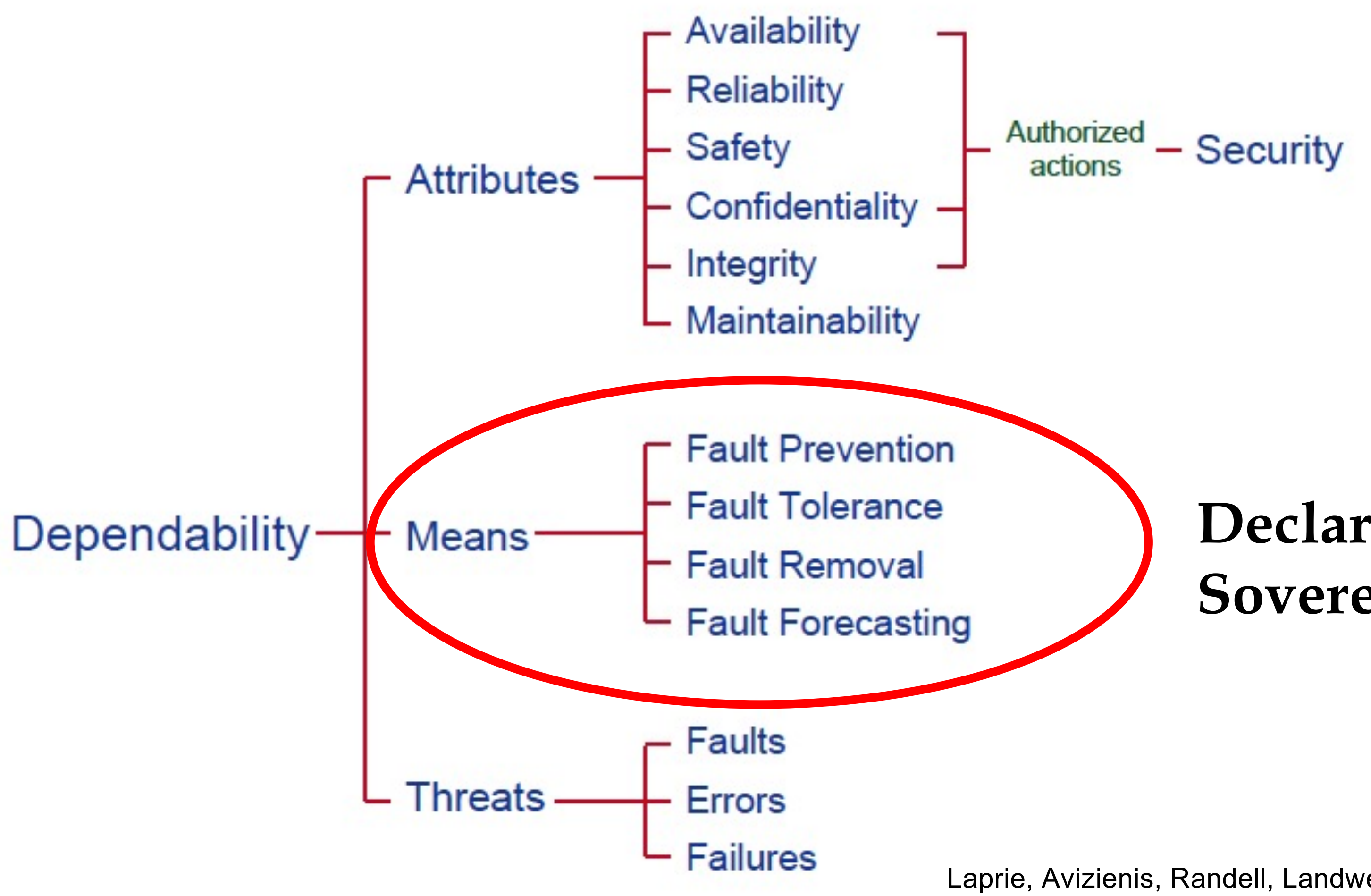
(What is expected from our research community?)

Extra slides

Cartography of our World

(Portuguese view circa 1500)

Cartography of (our) Dependability World



Declaration of Sovereignty ☺

Laprie, Avizienis, Randell, Landwehr, Simoncini,…