TRUSTWORTHY AUTONOMOUS SYSTEMS

Systems Engineering Challenges for Trustworthiness

Andre Bolles, Imke Hoppe, Axel Hahn, DLR, 2022-09-27

Why do we accept that people can drive and get their license after only 12 hours of driving lessons?

Why do we accept that people can drive and get their license after only 12 hours of driving lessons?

And why don't we accept an autonomous vehicle that has completed the same number of driving hours (in a simulator)?

Possible arguments

AV is not human

Possible arguments

AV is not human

Simulators are not reality

Possible arguments

AV is not human

Simulators are not reality

Cyber security

Possible arguments

AV is not human

Simulators are not reality

Sensor inaccuracy

Cyber security

Possible arguments

AV is not human

Simulators are not reality

Sensor inaccuracy

Explainability of algorithms

Cyber security

Possible arguments

AV is not human

Simulators are not reality

Sensor inaccuracy

Explainability of algorithms

Cyber security

Software bugs

Possible arguments

AV is not human

Simulators are not reality

Sensor inaccuracy

Explainability of algorithms

Cyber security

Software bugs

Electric errors

and many more..

But what is hidden behind these arguments?

TRUST

What makes a highly automated / autonomous system trustworthy?

Answers from the DLR Systems Engineering institute

comprehensible behavior

reliability

Runtime Assurances

Preparedness

Monitoring

Continuous (human) supervision

Criticality analysis

Certification Standards

Robustness

Adaptation to personal behavior

AI verification

less making me feel shocked

independent multiple protection

Reliable fall-back strategies

Knowledge

Easy take-over

Transparent decisions

Understandability

Transparency

Replicability

Proven robustness

Easy cooperation

Field experience

Explainability

Proven in use

testing

Integrity

making explainable decisions

Exhaustively tested

Formally proven

Error free

Predictable, comprehensible and explainable behavior

Verification

Proven in use

Interpretability

Trustworthy V&V Methods

thoroughly tested

Validation

Model validity

Realistic distribution of data

Approbiate Safety Level

“Trust […] is the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other part.”

(Mayer et al. 1995) summarizing (Kee and Knox 1970)

Is this achievable from a technical perspective only?

Or do we need more?

comprehensible behavior

reliability

Runtime Assurances

Preparedness

Monitoring

Continuous (human) supervision

Criticality analysis

Certification Standards

Robustness

Adaptation to personal behavior

AI verification

less making me feel shocked

independent multiple protection

Reliable fall-back strategies

Knowledge

Easy take-over

Transparent decisions

Understandability

Replicability

Proven robustness

Easy cooperation

Transparency

Explainability

Field experience

Proven in use

testing

Integrity

making explainable decisions

Exhaustively tested

Formally proven

Error free

Predictable, comprehensible and explainable behavior

Interpretability

Verification

Proven in use

Trustworthy V&V Methods

thoroughly tested

Validation

Model validity

Realistic distribution of data

Approbiate Safety Level

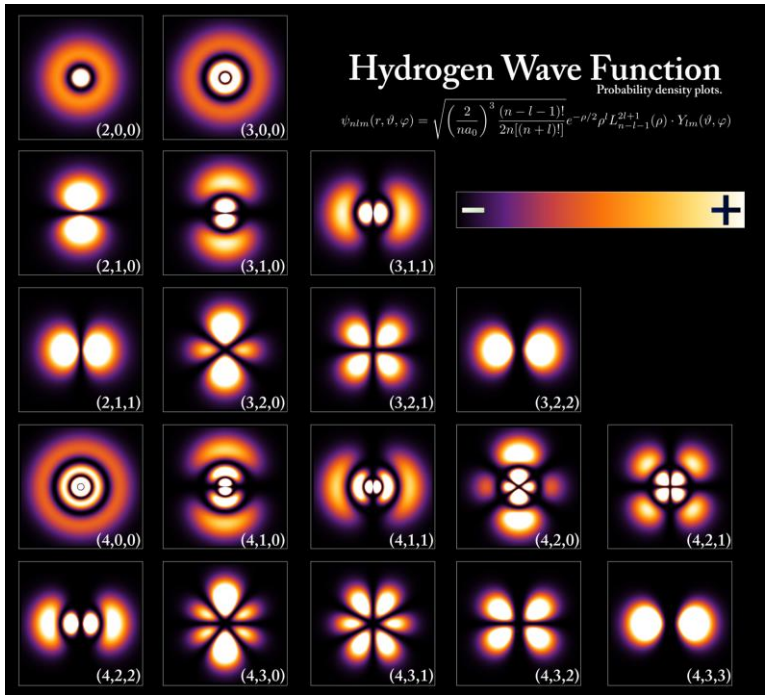
Obviously, we need more…

But where to get this?

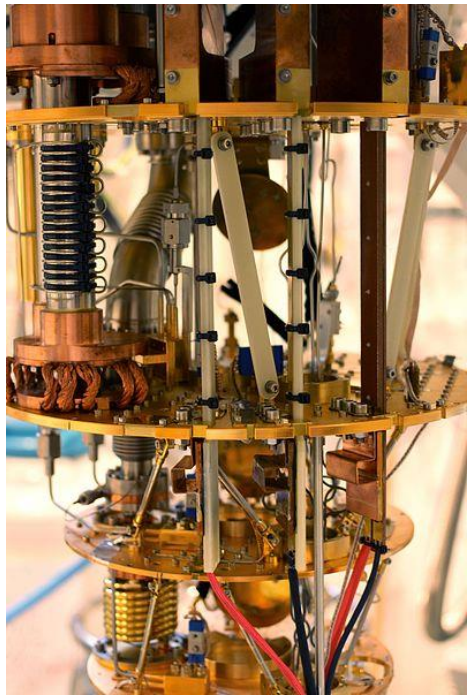
What did we do in the past, when we were in a dead end?

New Solutions for Sensors

Adopting physics for positioning by using inert sensors

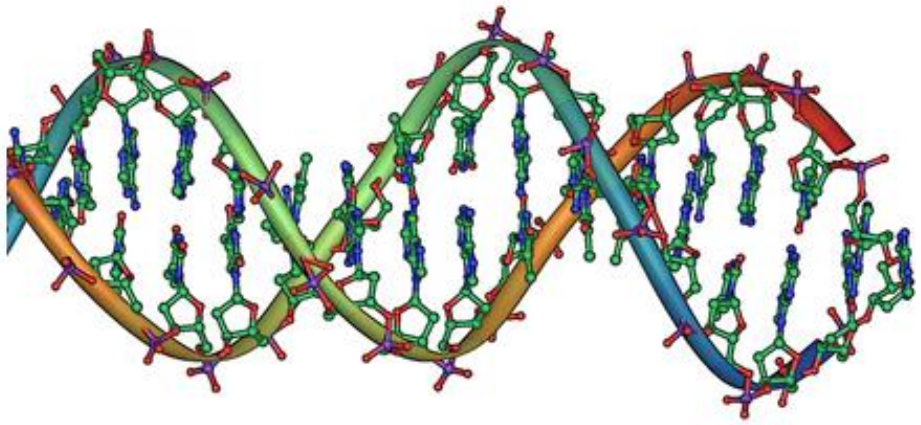


https://de.wikipedia.org/; public domain,



UCL Mathematical and Physical Sciences from London, UK
https://commons.wikimedia.org/wiki/File:Quantum_refrigerator_at_UCL_(17626619658).jpg
CC BY 2.0

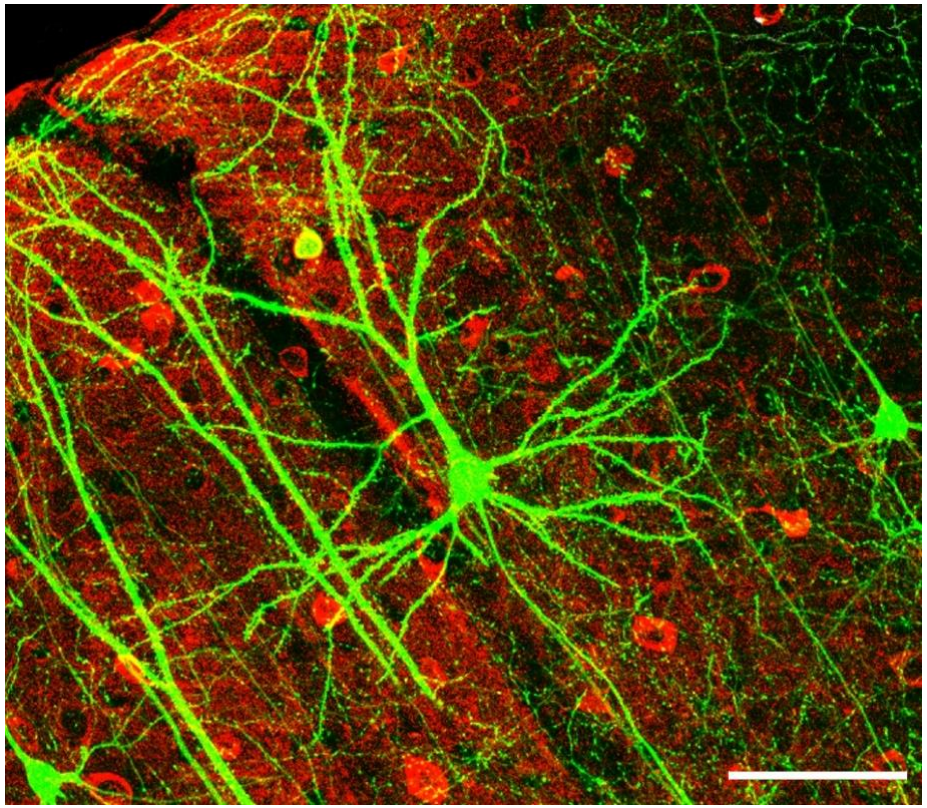
New Solutions for Parametrization
Adopting genetics for optimization



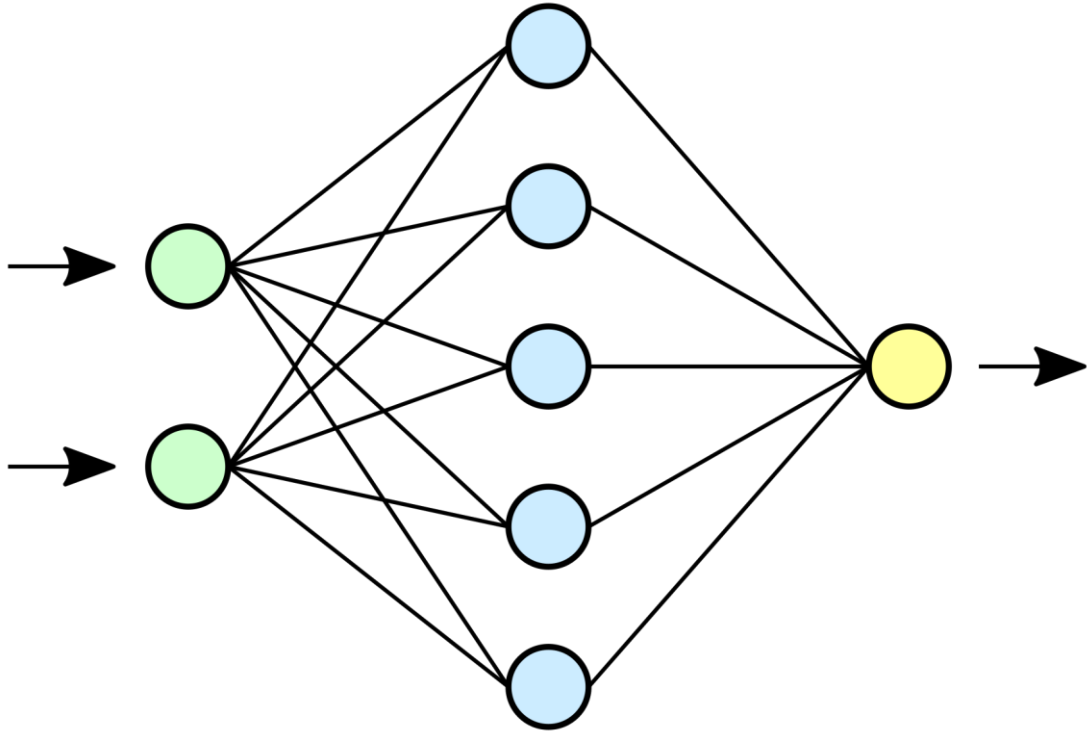
public domain, https://commons.wikimedia.org/wiki/File:DNA_double_helix_horizontal.png

public domain, en.wikipedia.org

New Solutions for System Functionality
Adopting biology by using neural networks for classification



https://de.wikipedia.org/wiki/Nervenzelle, CC BY 2.5



https://de.wikipedia.org/wiki/K%C3%BCnstliches_neuronales_Netz, CC BY 1.0

And what can we do in the future?

Autonomous Systems
Learning from human cooperation

CC BY 2.0
UNICEF Ukraine, commons.wikimedia.org

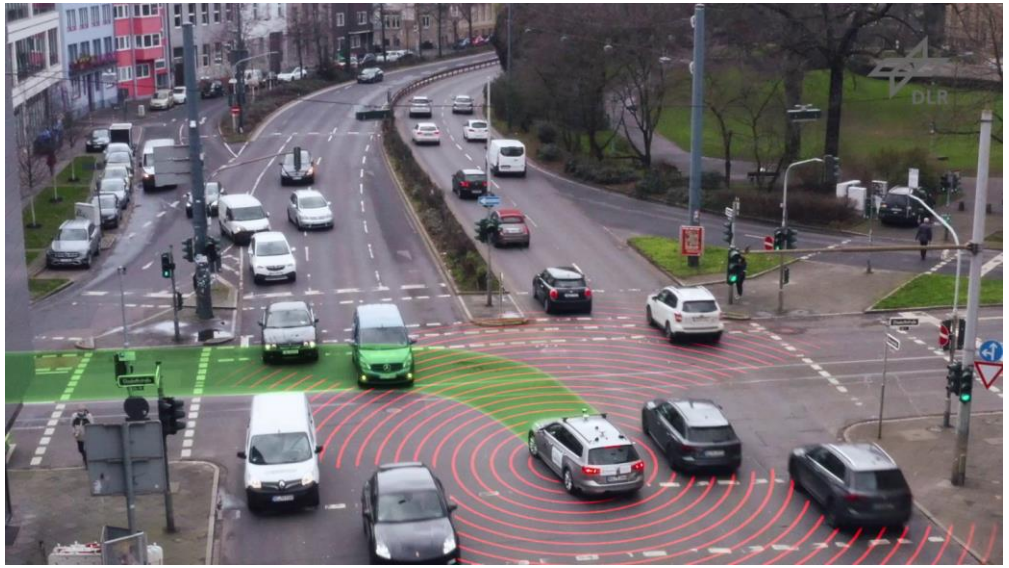
Intelligence of a social system: human interaction

public domain,
https://commons.wikimedia.org/wiki/File:Hong_Kong_Night_Skyline2.jpg

Intelligence of a technical system: Smart City

Siehe auch https://cmte.ieee.org/futuredirections/2017/08/15/tech-for-autonomous-systems-advanced-interaction-capabilities-iv/

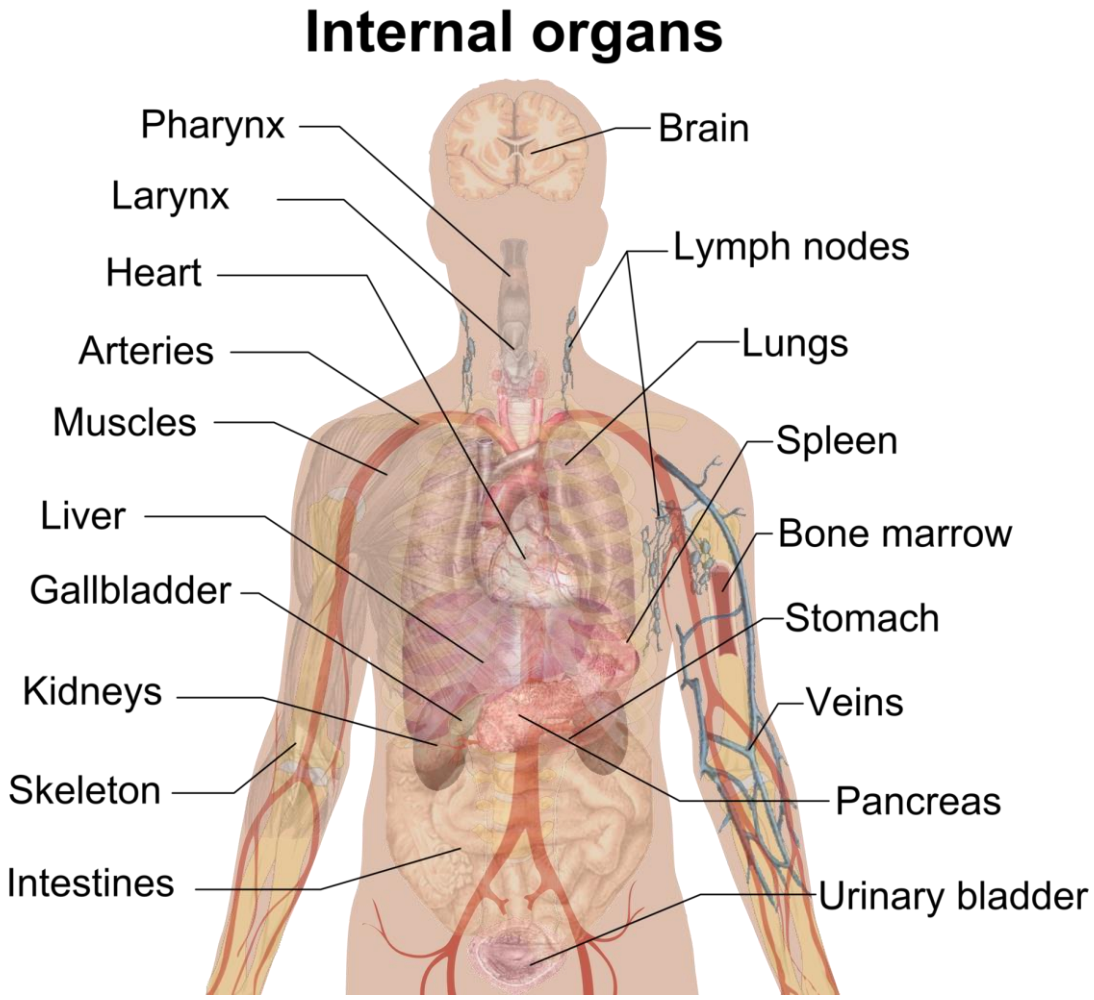
New System of Systems Design
Adopting social science for system management



DLR

New System Designs

Medical Science: adopting concepts of life for resilient complex systems



https://en.wikipedia.org/wiki/Organ_(biology), public domain

Dr. Karl-Heinz Hochhaus, CC-BY 3.0, https://commons.wikimedia.org/wiki/File:2012_09_03_Br%C3%BCckenfahrstand_Voraus_DSCI73357.JPG

New System Behaviour
Adopting complexity theory for self guiding sytems

CC BY 2.0;
https://commons.wikimedia.org/wiki/File:Fish_swarm_through_the_kelp_forest.jpg

CC BY 2.0
https://commons.wikimedia.org/wiki/File:Driving_Cars_in_a_Traffic_Jam.jpg

Remembering what we learned so far…

*not automated

Maybe we should consider autonomous* systems as „living“ entities with an integrated role in our societies…

*not automated

…covering not only trust but also other aspects of socialisation…

*not automated

…and address this situation from several disciplines…

*not automated

…implicating also changes in systems engineering…

*not automated

Autonomics
System Engineering Challenges

New Technologies and Solutions for System Functionality

DLR, CC BY NC ND 3.0

Open your mind:

- Find new solutions
- Adopt your thinking
- Identifiy and transfer technologies

New Designs and Systems Architectures

Leo Luijten,
https://commons.wikimedia.org/wiki/File:Architectur_in_Rotterdam_(39938175532).jpg
CC BY 2.0

Leave the beaten tracks:

- Out of the box thinking
- new system architecure patterns

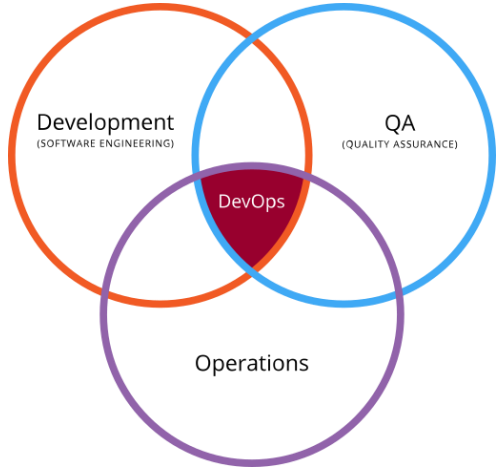
New Engineering Methods and Tools

DLR, CC-BY 3.0

New Thinking:

- Find and apply new aproaches for design and analysis
- Handle new dimensions of system complexity

New Engineering Methodologies



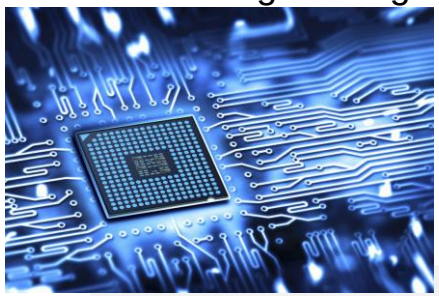
https://commons.wikimedia.org/wiki/File:Devops.svg,
CC BY 3.0

New Cooperation:

- Systematically explore the design space
- Rethink the life cycle
- Transdisceplinary teams
- New roles in engineering

Evolution of Engineering*

Electronic Engineering



https://commons.wikimedia.org/wiki/File:Electronics-projects.jpg, CC BY-SA 4.0

Math / Informatics

Yulia Meshkova, https://commons.wikimedia.org/wiki/File:The_board.jpg; CC BY 4.0

Digitalization

DLR

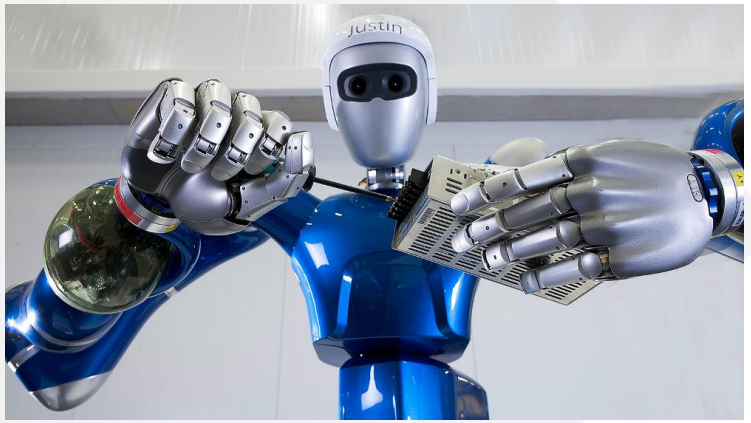
Mechatronics

Steve Jurvetson, https://commons.wikimedia.org/wiki/File:Tesla_auto_bots.jpg; CC BY 2.0

Systems Engineering

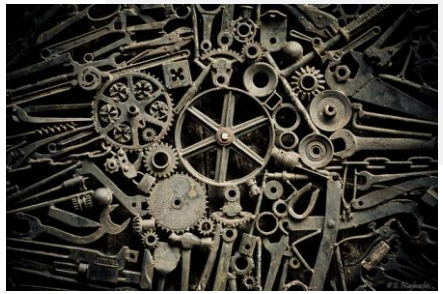
DLR, CC BY-NC-ND 3.0

Autonomics*



DLR (CC-BY 3.0)

Mechanical Engineering



"Old Cogs" von Emmanuel Huybrechts. Lizenz: CC BY 2.0

Transdisciplinary Research

https://commons.wikimedia.org/wiki/File:L%C3%A1pices_de_colores_01.jpg, CC BY 2.0, Manuel Martín Vicente

*idea originally proposed by Peter Liggesmeyer on https://link.springer.com/content/pdf/10.1007/s00287-017-1046-1.pdf

Imprint

Topic:	**Trustworthy Autonomous Systems**
	Systems Engineering Challenges for Trustworthiness
Date:	2022-09-27
Authors:	André Bolles, Imke Hoppe, Axel Hahn
Institute:	DLR Institute of Systems Engineering for Future Mobility
Credits:	mentioned on the images

Literature

- (Mayer et al. 1995): Mayer, Roger C.; Davis, James H.; Schoorman, F. David. An Integrative Model of Organizational Trust. The Academy of Management Review 20 (3), p. 709 – 734
- (Kee and Knox 1970): Kee, Herbert W.; Knox, Robert E. Conceptual and methodological considerations in the study of trust and suspicion. Journal of conflict resolution, 14 (3), p. 357 – 366
- (TÜV 2022) TÜV Mobility Studie 2022, Zukunft der Mobilität nachhaltig, digital, sicher