TRUSTWORTHY AUTONOMOUS SYSTEMS
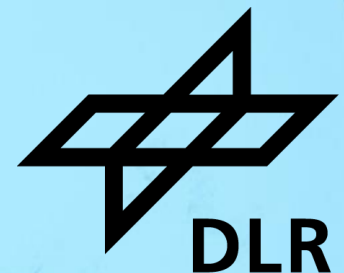
Systems Engineering Challenges for Trustworthiness

(revised version of the original presentation)

DLR

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
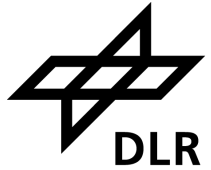
Cyber security

Software bugs

Electric errors

and many more..

But what is hidden behind these arguments?

TRUST

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

What makes a highly automated / autonomous systems trustworthy?

Answers from the DLR Institute of Systems Engineering for Future Mobility

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

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

Interpretability

Predictable, comprehensible and explainable behavior

Verification

Proven in use

Trustworthy v&v methods

Thoroughly tested

Validation

Model validity

Aprobate safety sevel

Realistic distribution of data

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." [1, p. 712]

Is this achievable from a technical perspective only?

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

Or do we need more?

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

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

Interpretability

Predictable, comprehensible and explainable behavior

Verification

Proven in use

Trustworthy v&v methods

Thoroughly tested

Validation

Model validity
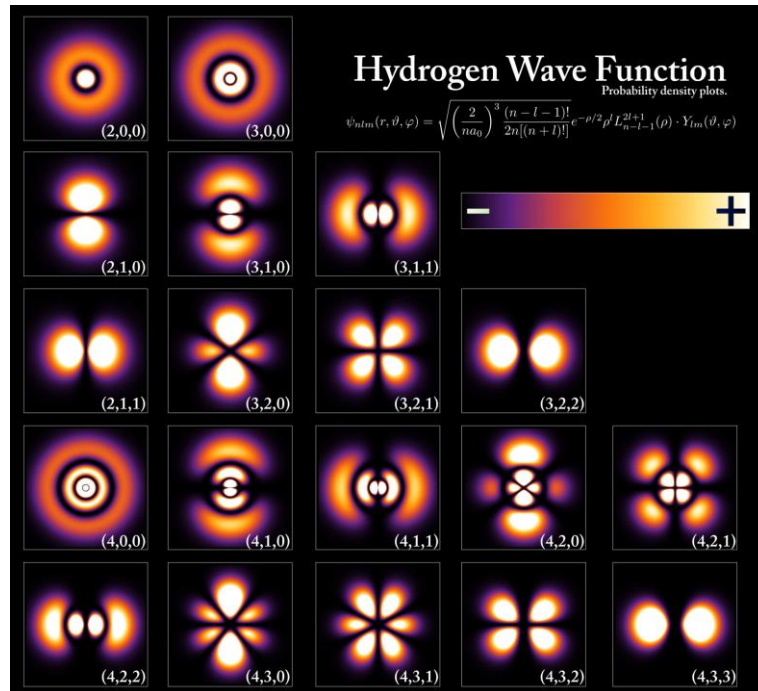
Aprobate safety sevel

Realistic distribution of data

DLR

Obviously, we need more…

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

But where to get this?

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

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

Title (from source): "Hydrogen Density Plots for n up to 4. Wavefunctions of the electron in a hydrogen atom at different energy levels"
By: Wikipedia user PoorLeno referencing Forinash, Kyle. Hydrogen W Simulation. Indiana University Southeast. Retrieved on 2008-12-18. Tokita, Sumio; Sugiyama, Takao; Noguchi, Fumio; Fujii, Hidehiko; Kobayashi, Hidehiko (2006). "An Attempt to Construct an Isosurface Having Symmetry Elements". Journal of Computer Chemistry, Japan 5 (3): 159–164. DOI:10.2477/jccj.5.159,
Source: https://commons.wikimedia.org/wiki/File:Hydrogen_Density_Plots.png
Last visit: 2022-12-06
License: public domain,
Changes: no changes made
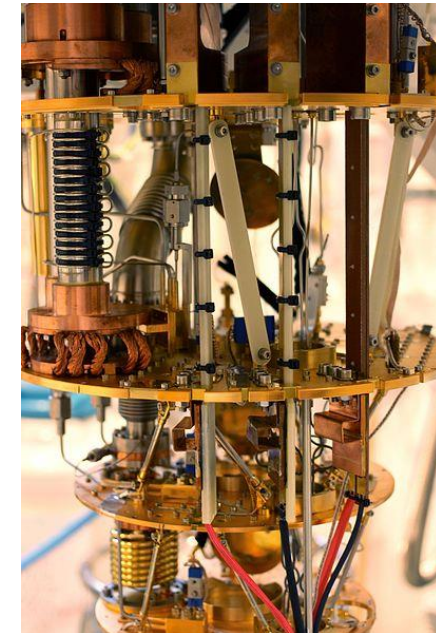
Title (from source): "Quantum refrigerator at UCL"
By: O. Usher (UCL MAPS), UCL Mathematical and Physical Sciences from London, UK
Source: https://commons.wikimedia.org/wiki/File:Quantum_refrigerator_at_UCL_(17626619658).jpg
Last visit: 2022-12-06
License: CC BY 2.0,
Changes: no changes made

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

New Solutions for Parametrization
## Adopting genetics for optimization
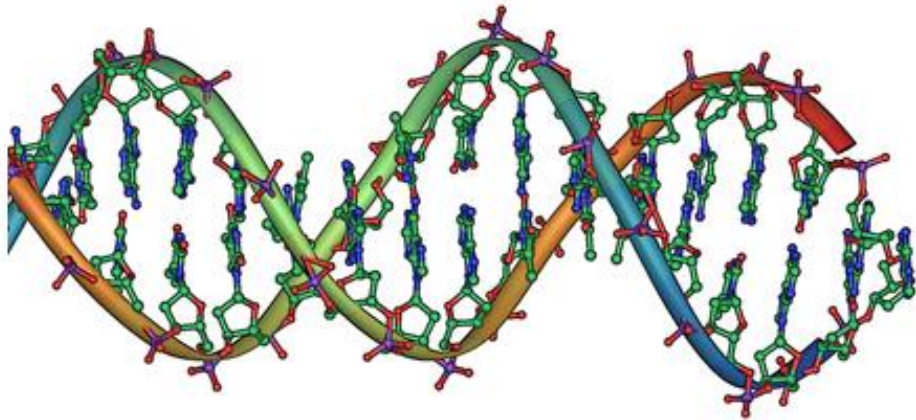
Title (from source): „DNA double helix horizontal"
By: Jerome Walker
Source: https://commons.wikimedia.org/wiki/File:DNA_double_helix_horizontal.png
Last visit: 2022-12-06
License: public domain
Changes: no changes made

Title (from source): „St 5-xband-antenna"
By: NASA
Source: https://commons.wikimedia.org/wiki/File:St_5-xband-antenna.jpg,
Last visit: 2022-11-06
License: public domain
Changes no changes made

Andre Bolles, Imke Hoppe, Axel Hahn, DLR, 2022-09-27
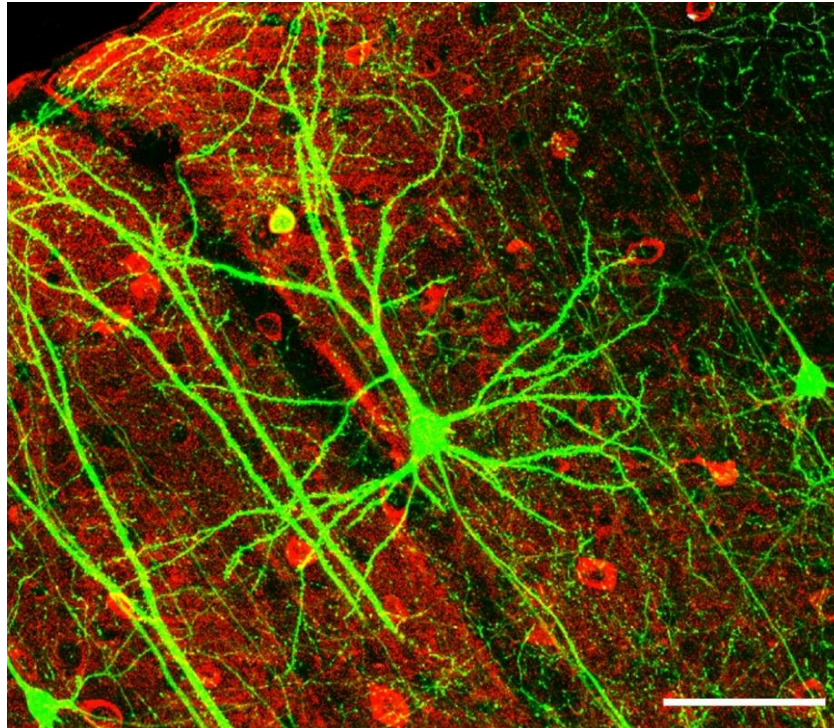
Title (from source): "Coronal section containing the chronically imaged pyramidal neuron "dow" (visualized by green GFP) does not stain for GABA (visualized by antibody staining in red). Confocal image stack, overlay of GFP and GABA channels. Scale bar: 100 µm"
Authors: Wei-Chung Allen Lee, Hayden Huang, Guoping Feng, Joshua R. Sanes, Emery N. Brown, Peter T. So, Elly Nedivi
Source: https://commons.wikimedia.org/wiki/File:PLoSBiol4.e126.Fig6fNeuron.jpg
Last visit: 2022-12-06
License: CC BY 2.5
Changes: no changes made
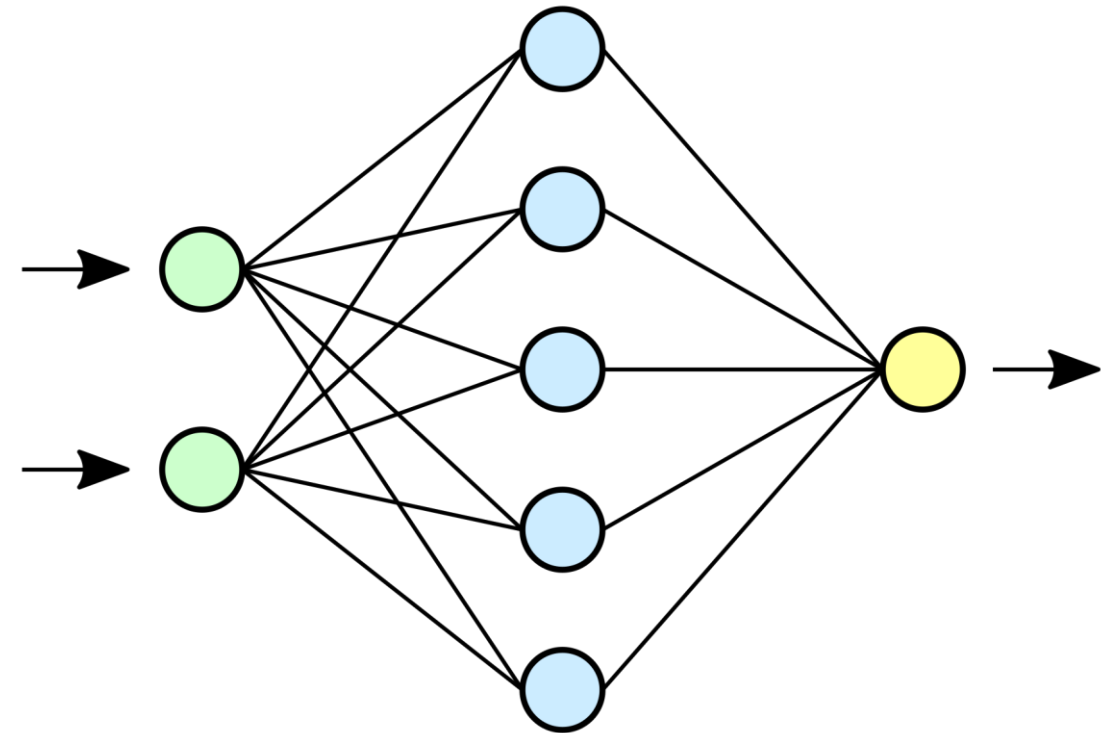
Title (from source): "A simplified view of an artifical neural network"
Authors: Wikimedia users Dake and Mysid
Source: https://commons.wikimedia.org/wiki/File:Neural_network.svg,
Last visit: 2022-12-06
License: CC BY 1.0
Changes: no changes made,

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

And what can we do in the future?

Autonomous Systems
Learning from human cooperation

Title (from source): -
By: UNICEF Ukraine
Source: https://commons.wikimedia.org/wiki/File:UNICEF_DSC03709_(21074874624).jpg
Last visit: 2022-12-06
License: CC BY 2.0
Changes: no changes made

Title (from source): „Skyline of Victoria Habour[sic!]"
By: Wikimedia user Scrollock
Source: https://commons.wikimedia.org/wiki/File:Hong_Kong_Night_Skyline2.jpg
Last visit: 2022-12-06
License: public domain
Changes: added a transparent overlay for illustrating digitalization, cut off some parts of the original picture at the left and right side
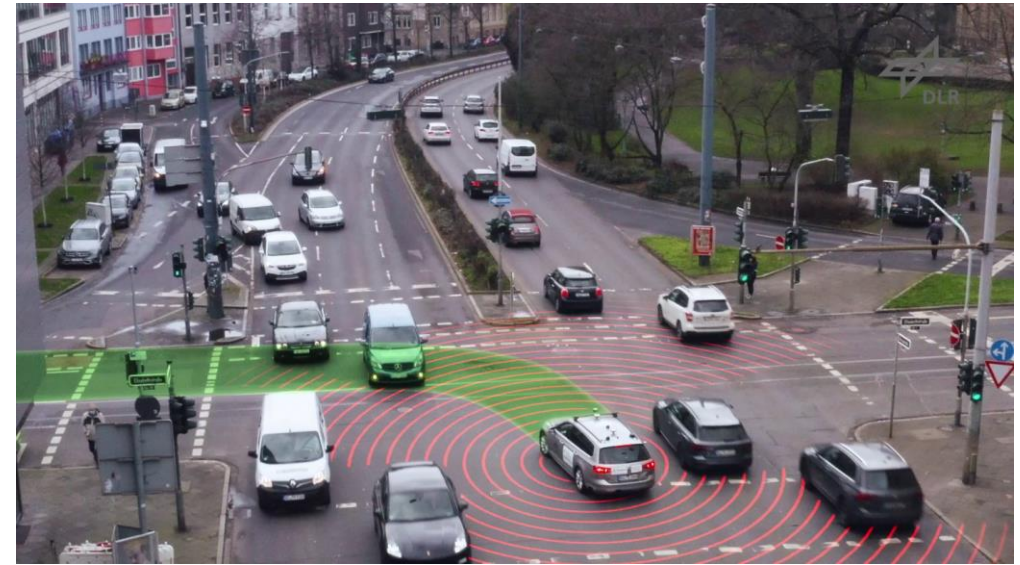
Intelligence of a social system: human interaction   **Intelligence of a technical system: Smart City**

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

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

New System of Systems Design
Adopting social science for system management

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

Internal organs

Pharynx — Brain
Larynx
Heart — Lymph nodes
Arteries — Lungs
Muscles — Spleen
Liver — Bone marrow
Gallbladder — Stomach
Kidneys — Veins
Skeleton — Pancreas
Intestines — Urinary bladder

Credit as proposed in the source link:
Häggström, Mikael (2014). "Medical gallery of Mikael Häggström 2014".
WikiJournal of Medicine **1** (2). DOI:10.15347/wjm/2014.008 . ISSN 2002-4436. Public Domain.

Source: https://commons.wikimedia.org/wiki/File:Internal_organs.png
Last visit: 2022-12-06
License: public domain
Changes: no changes made

Title (from source): „Moderner Fahrstand auf der Brücke"
By: Dr. Karl-Heinz Hochhaus
Source: https://commons.wikimedia.org/wiki/File:2012_09_03_Br%C3%BCckenfahrstand_Voraus_DSCI73357.JPG,
Last visit: 2022-12-06
Licence: CC-BY 3.0
Changes: no changes made

New System Behaviour
Adopting complexity theory for self guiding sytems

Title (from source): „Fish swarm through the kelp forest"
By: Wikimedia user oliver.dodd
Source: https://commons.wikimedia.org/wiki/File:Fish_swarm_through_the_kelp_forest.jpg
Last visit: 2022-12-06
License: CC BY 2.0
Changes: no changes made
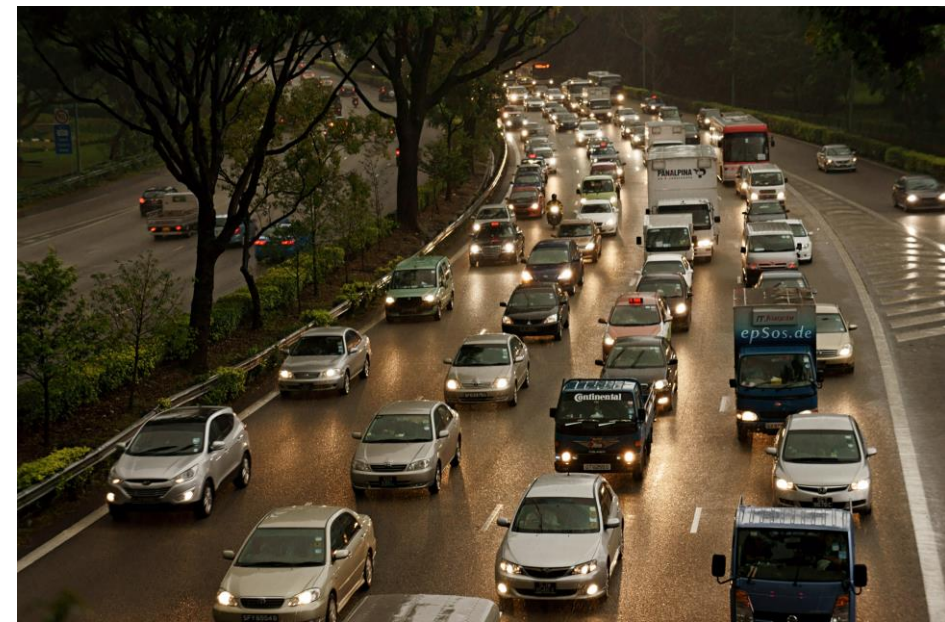
Title (from source): „Driving Cars in a Traffic Jam"
By: epSos.de
Source: https://commons.wikimedia.org/wiki/File:Driving_Cars_in_a_Traffic_Jam.jpg,
Last visit: 2022-10-06
License: CC BY 2.0
Changes: no changes made

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

Remembering what we learned so far…

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

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

*not automated

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

…and address this situation from several disciplines…

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

…implicating also changes in systems engineering…

System Engineering Challenges

New Technologies and Solutions for System Functionality

Open your mind:
- Find new solutions
- Adopt your thinking
- Identifiy and transfer technologies

New Designs and Systems Architectures

Leave the beaten tracks:
- Out of the box thinking
- new system architecure patterns

New Engineering Methods and Tools

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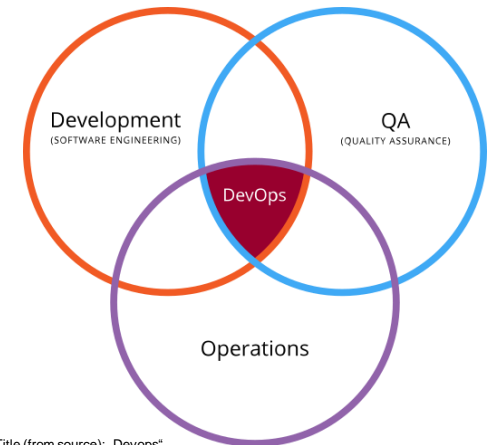- Find and apply new aproaches for design and analysis
- Handle new dimensions of system complexity

New Engineering Methodologies

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

Evolution of Engineering*

Electronic Engineering

Figure 37.1, credits see end of presentation (slide 40)

Math / Informatics

Figure 37.3, credits see end of presentation (slide 40)

Digitalization
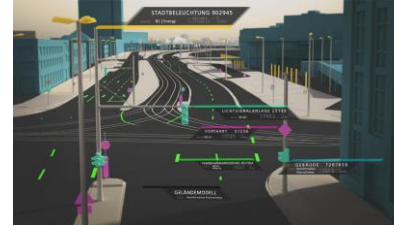
Figure 37.5, credits see end of presentation (slide 40)

Autonomics*

Figure 37.8, credits see end of presentation (slide 40)

Systems Engineering

Figure 37.6, credits see end of presentation (slide 40)

Mechatronics

Figure 37.4, credits see end of presentation (slide 40)

Mechanical Engineering

Figure 37.2, credits see end of presentation (slide 40)

Transdisciplinary Research

Figure 37.7, credits see end of presentation (slide 40)

*idea originally proposed by Peter Liggesmeyer on https://link.springer.com/content/pdf/10.1007/s00287-017-1046-1.pdf, last visit: 2022-10-06

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

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: see figures and slide 40

References (Citation Style IEEE)

[1] R. C. Mayer, J. H. Davis, and F. D. Schoorman,
"An Integrative Model of Organizational Trust," The Academy of Management Review,
vol. 20, no. 3, pp. 709–734, 1995.

Credits

Figure 37.1
Title (from source): "Circuito integrado de un escáner Artec"
By: Wiikmedia user Amfeli
Source: https://commons.wikimedia.org/wiki/File:Circuito_integrado.jpg
Last visit: 2022-12-06
License: CC BY 4.0
Changes: no changes made

Figure 37.2
Title (from source): "Old Cogs"
By: Emmanuel Huybrechts
Source: https://commons.wikimedia.org/wiki/File:Old_Cogs_(5084228263).jpg
Last visit: 2022-12-06
License: CC BY 2.0
Changes: no changes made

Figure 37.3
Title (from source): „The lecture notes at the board"
By: Yulia Meshkova
Source: https://commons.wikimedia.org/wiki/File:The_board.jpg
Last visit: 2022-12-06
License: CC BY 4.0
Changes: no changes made

Figure 37.4
Title (from source): „Tesla auto bots"
By: Steve Jurvetson
Source: https://commons.wikimedia.org/wiki/File:Tesla_auto_bots.jpg
Last visit: 2022-12-06
License: CC BY 2.0
Changes: no changes made

Figure 37.5
Title (from source): „Verkehrsrelevante Elemente im Digitalen Atlas"
By: Deutsches Zentrum für Luft- und Raumfahrt e. V. (DLR)
Source: https://www.dlr.de/content/de/artikel/digitalisierung/projekt-digitaler-atlas.html
Last visit: 2022-12-06
License: © DLR
Changes: no changes made

Figure 37.6
Title (from Source): „Forschungsinfrastruktur NGT FuN"
By: Deutsches Zentrum für Luft- und Raumfahrt e. V. (DLR)
Source: https://www.dlr.de/content/de/artikel/news/2022/03/20220913_hightech-fahrwerk-fuer-den-zug-der-zukunft.html
Last visit: 2022-12-06
License: CC BY-NC-ND 3.0
no changes made, last visit: 2022-10-06

Figure 37.7
Title (from source): „Lápices de colores"
By: Manuel Martín Vicente
Source: https://commons.wikimedia.org/wiki/File:L%C3%A1pices_de_colores_01.jpg
Last visit: 2022-12-06
License: CC BY 2.0
Changes: no changes made

Figure 37.8
Title (from source): „Rollin' Justin kehrt die Scherben einer zerbrochenen Tasse"
By: Deutsches Zentrum für Luft- und Raumfahrt e. V. (DLR)
Source (article): https://www.dlr.de/rm/desktopdefault.aspx/tabid-11427/#gallery/27296
Source (file):
https://www.dlr.de/rm/Portaldata/52/Resources/Roboter_und_Systeme/Justin/Rollin_Justin/scherben_kehren/rollin_justin_sweeping_orig.jpg,
Last Visit (both): 2022-12-06
License: CC-BY 3.0,
Changes: no changes made