2025 International Conference on Cyber Security and Resilience (IEEE CSR'25)
Chania, Crete, Greece, August 4-6, 2025

Digital Twins for Trustworthy Autonomy

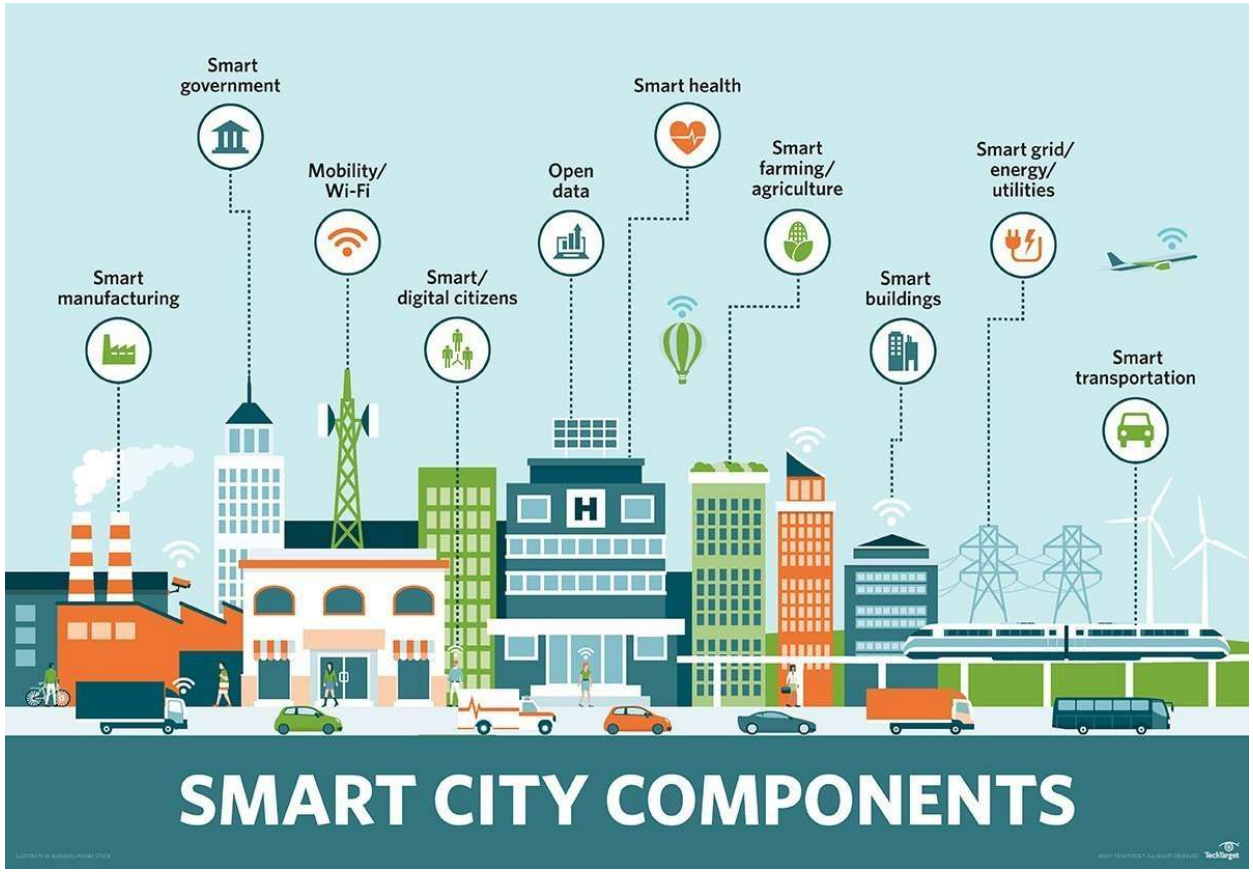
Prof. Francesco Flammini, Ph.D.

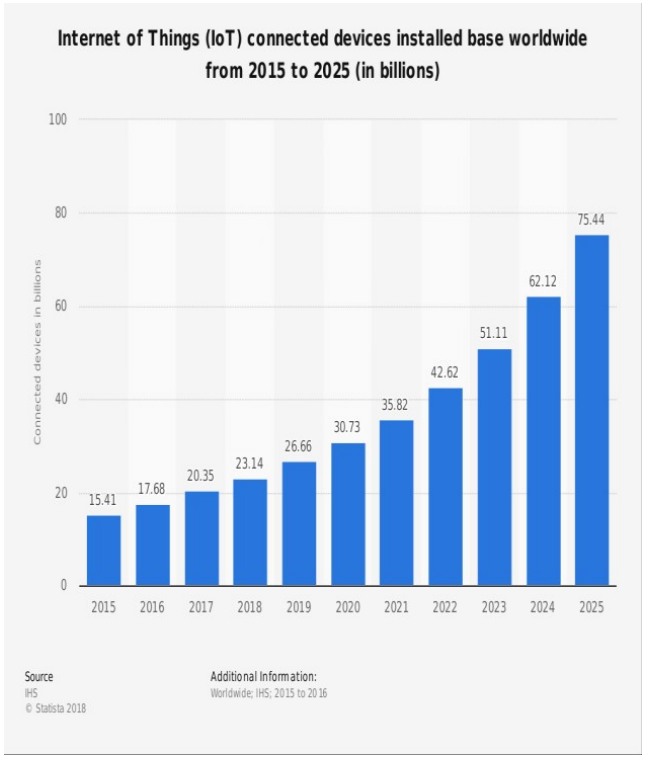
IDSIA USI-SUPSI, and

University of Florence, DIMAI

francesco.flammini@ieee.org

Smart-X applications and the Internet of everything





GRAND CHALLENGES

- Need for designing cyber-physical systems (CPS) that are:
  - Ubiquitous and pervasive
  - Smart, intelligent and autonomous
  - Reliable, safe and secure

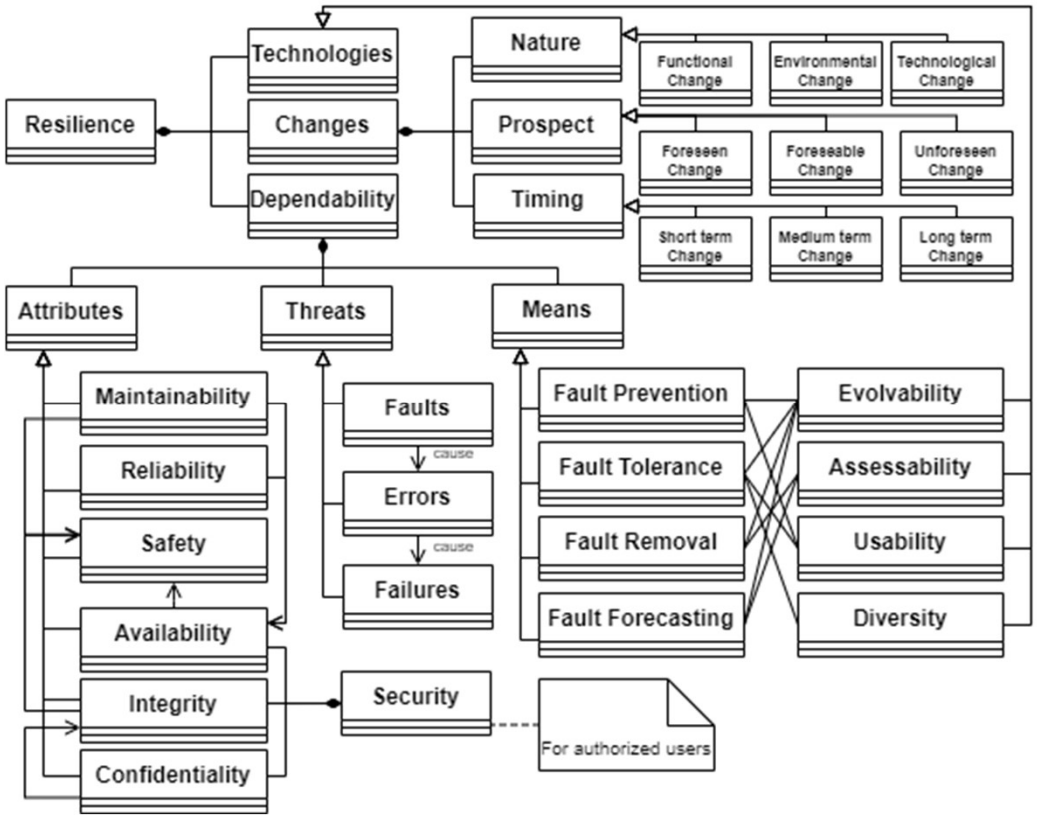
- How to manage complexity, heterogeneity, and uncertainties?

F. Flammini et al. (2019). Complex, Resilient and Smart Systems. In: (edited by): F. Flammini, Resilience of Cyber-Physical Systems: From Risk Modeling to Threat Counteraction. ADVANCED SCIENCES AND TECHNOLOGIES FOR SECURITY APPLICATIONS, p. 3-24, BERLIN HEIDELBERG:Springer-Verlag, ISBN 978-3-319-95597-1, ISSN: 1613-5113

What is resilience?

- “The persistence of service delivery that can justifiably be trusted, when facing **changes**.”
- “The persistence of the avoidance of failures that are unacceptably frequent or severe, when facing **changes**.”
- “The persistence of dependability when facing **changes**.”

(Jean-Claude Laprie, 2008)



F. Flammini, C. Alcaraz, E. Bellini, S. Marrone, J. Lopez and A. Bondavalli, "Towards Trustworthy Autonomous Systems: Taxonomies and Future Perspectives," in *IEEE Transactions on Emerging Topics in Computing*, doi: 10.1109/TETC.2022.3227113.

What is RISK?

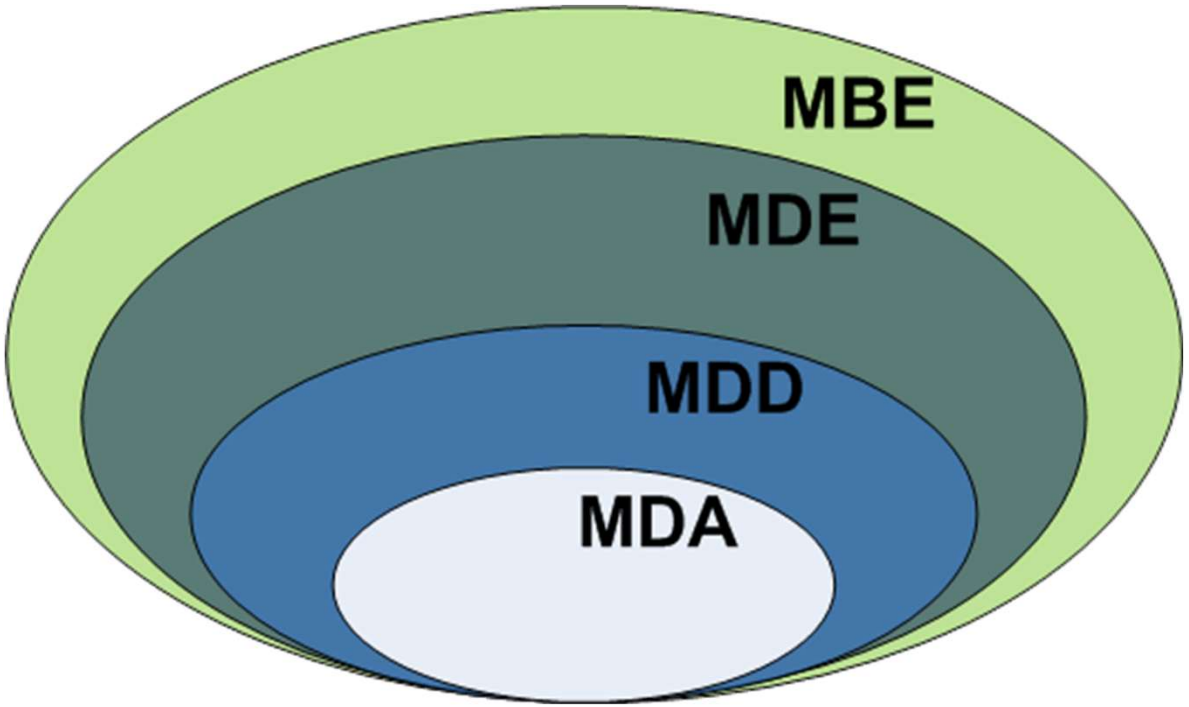
$$R = P \times V \times D$$

Several quantitative non-functional requirements, e.g.:

- MTBF > 500.000 h
- THR < 10^{-9} hazards / h

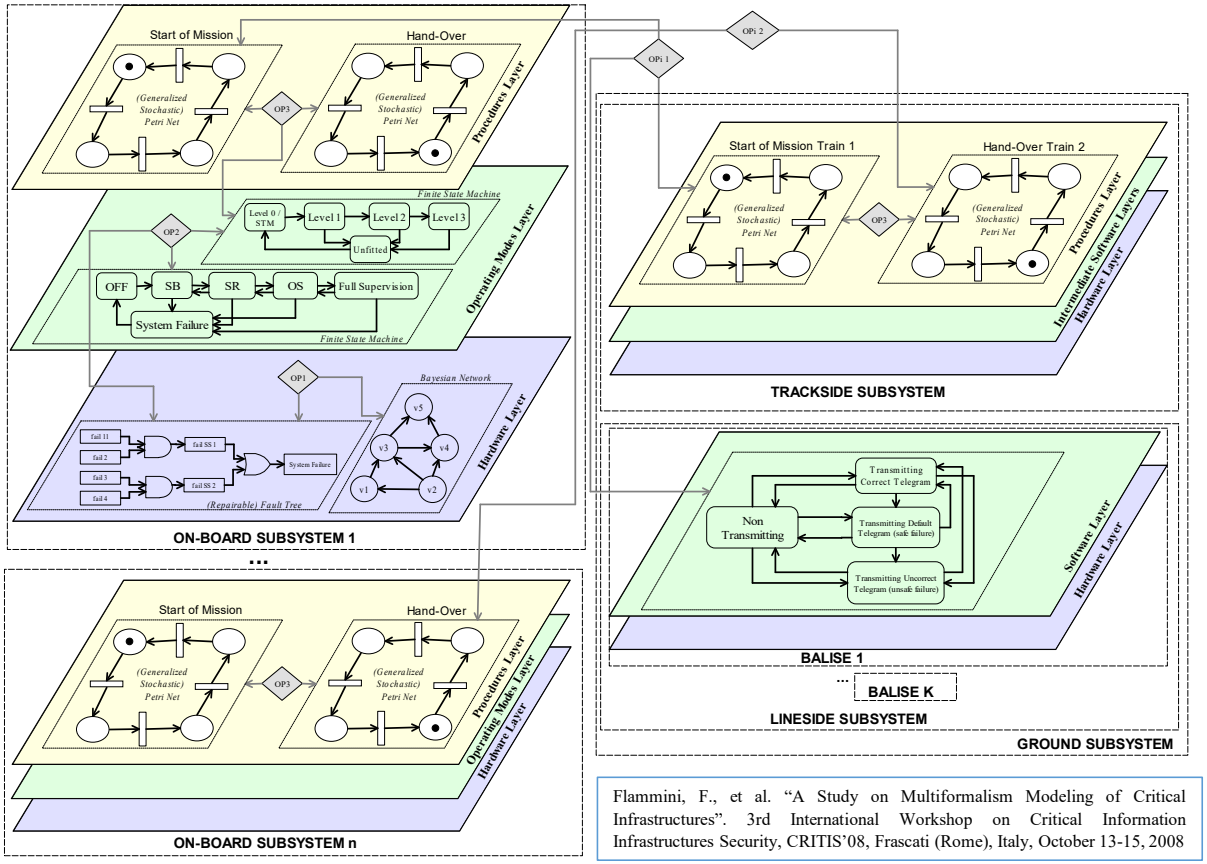
Flammini F., Gentile U., Marrone S., Nardone R., Vittorini V. (2014) A Petri Net Pattern-Oriented Approach for the Design of Physical Protection Systems. In: Bondavalli A., Di Giandomenico F. (eds) Computer Safety, Reliability, and Security. SAFECOMP 2014. Lecture Notes in Computer Science, vol 8666. Springer, Cham. https://doi.org/10.1007/978-3-319-10506-2_16

Model-based engineering



https://modeling-languages.com/clarifying-concepts-mbe-vs-mde-vs-mdd-vs-mda/

Multi-paradigm modeling



Flammini, F., et al. "A Study on Multiformalism Modeling of Critical Infrastructures". 3rd International Workshop on Critical Information Infrastructures Security, CRITIS'08, Frascati (Rome), Italy, October 13-15, 2008

Certification


- Could be worse.
- How?
Could be raining.

Self-x technologies

Level 0	Level 1	Level 2
Warnings	Cooperation	Partial Autonomy
Advanced Driver Assistance Systems		
Level 3	**Level 4**	**Level 5**
Conditional Autonomy	High Autonomy	Full Autonomy
Autonomous Vehicles		

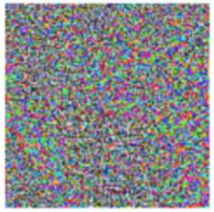
N. Rajabli, F. Flammini, R. Nardone and V. Vittorini, "Software Verification and Validation of Safe Autonomous Cars: A Systematic Literature Review," in *IEEE Access*, vol. 9, pp. 4797-4819, 2021, doi: 10.1109/ACCESS.2020.3048047.

Adversarial attacks to AI

$\boldsymbol{x}$

"panda"
57.7% confidence

$+ .007 \times$



$\text{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y))$

"nematode"
8.2% confidence

$=$

$\boldsymbol{x} + \epsilon\text{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y))$

"gibbon"
99.3 % confidence

Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., and Fergus, R. **Intriguing properties of neural networks**. ICLR (2013).

AI for safety critical systems

Growing concerns about safety-critical settings with AI

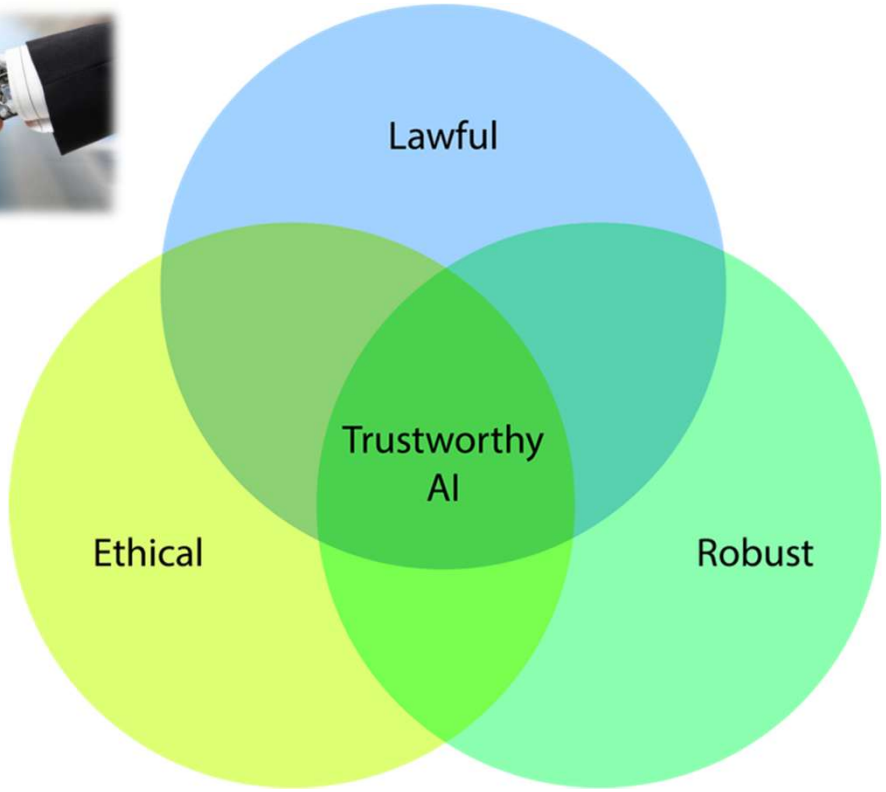
Autonomous cars that deploy AI model for traffic signs recognition

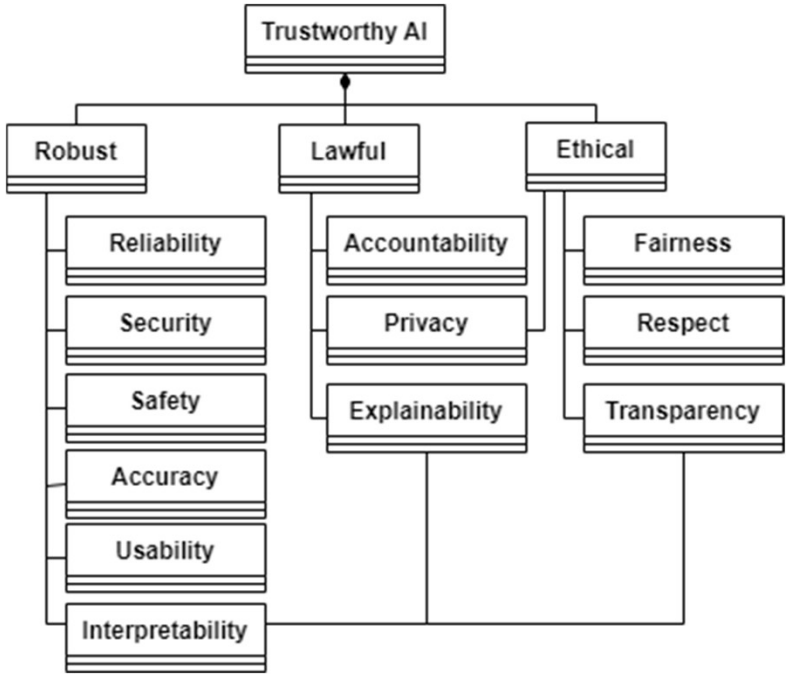
But with adversarial examples…

- https://www.eetimes.eu/2019/02/20/ai-tradeoff-accuracy-or-robustness/
- https://dorsa.fyi/cs521/

(IBM Research)

Trustworthy artificial intelligence



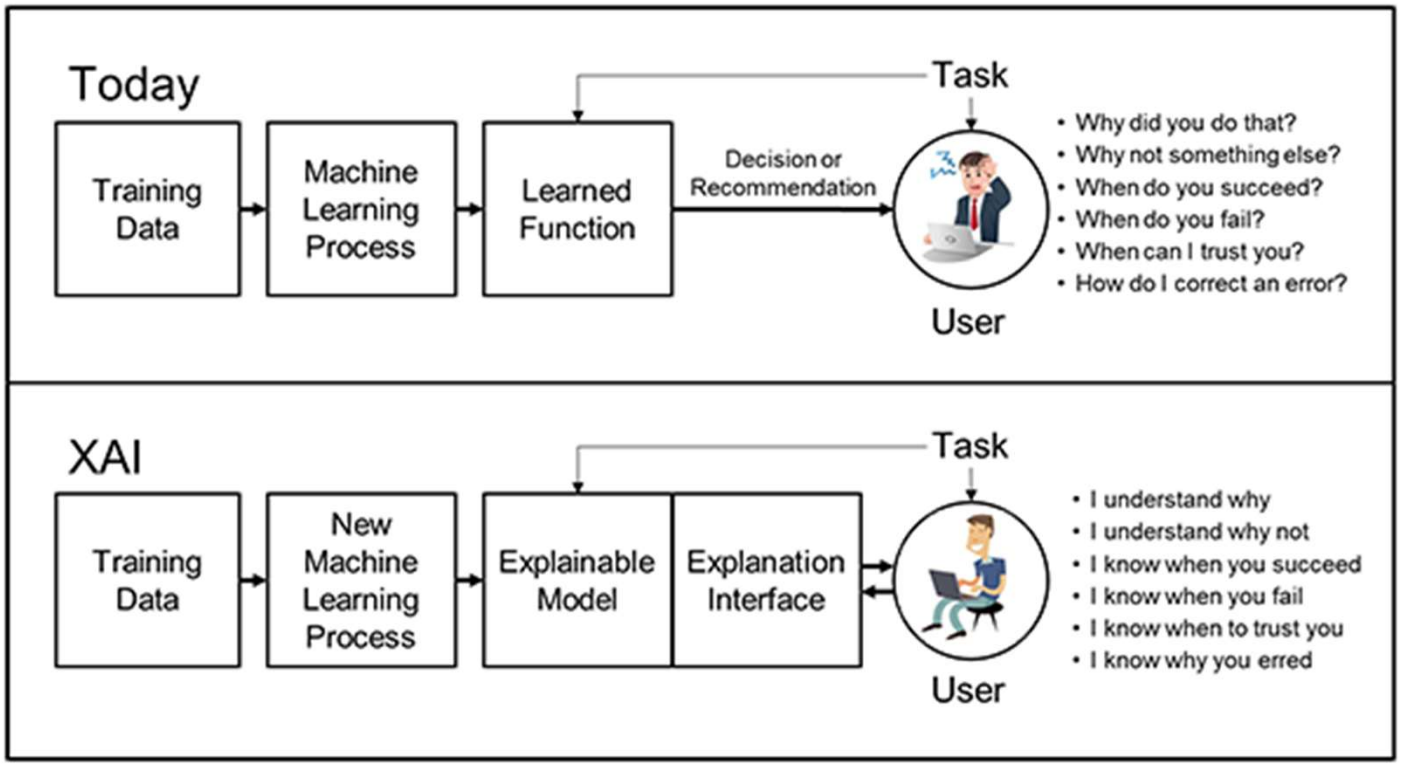


F. Flammini, C. Alcaraz, E. Bellini, S. Marrone, J. Lopez and A. Bondavalli, "Towards Trustworthy Autonomous Systems: Taxonomies and Future Perspectives," in *IEEE Transactions on Emerging Topics in Computing*, doi: 10.1109/TETC.2022.3227113.

IEEE

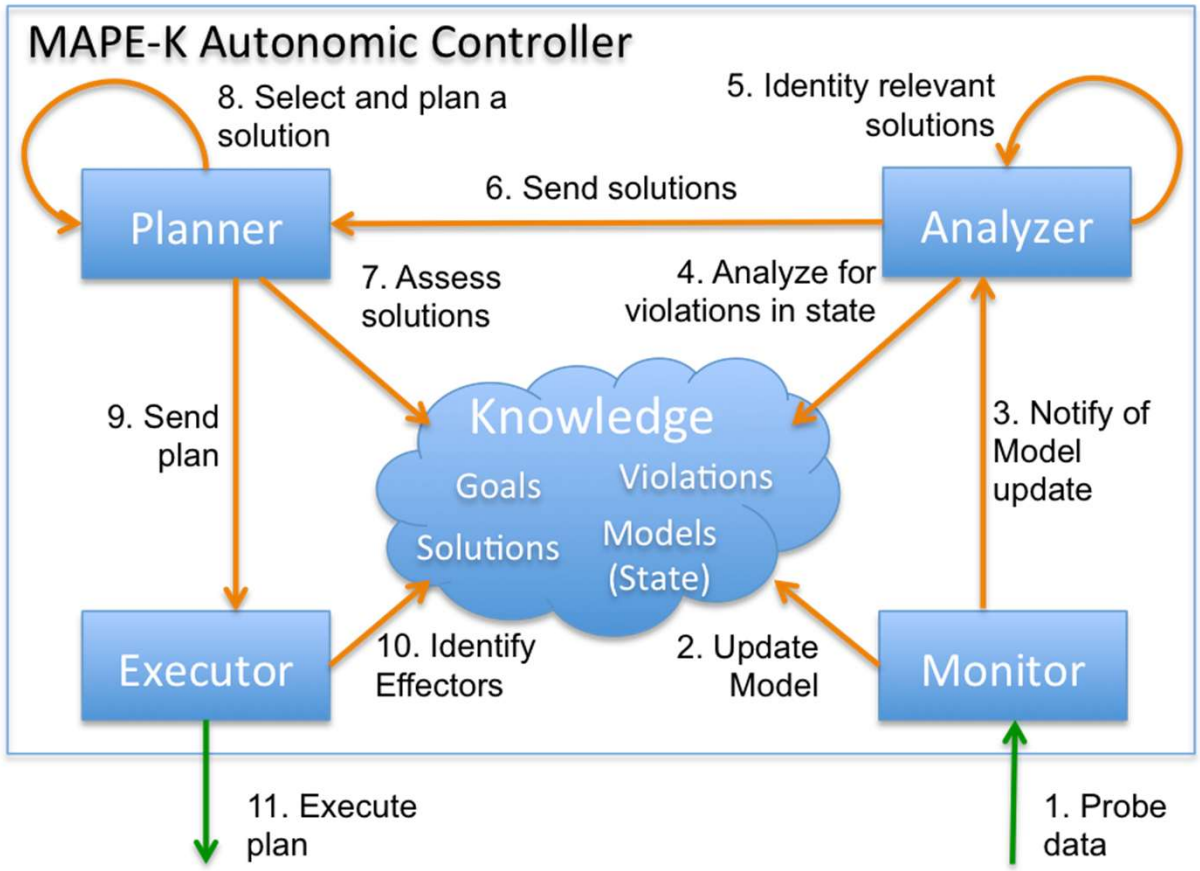
https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

Explainable artificial intelligence



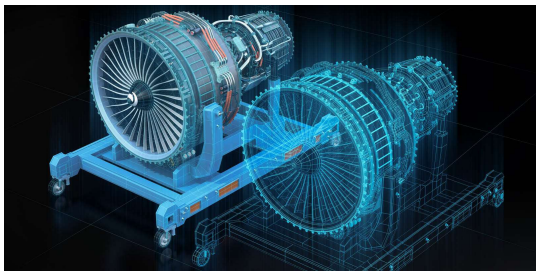
https://www.darpa.mil/program/explainable-artificial-intelligence

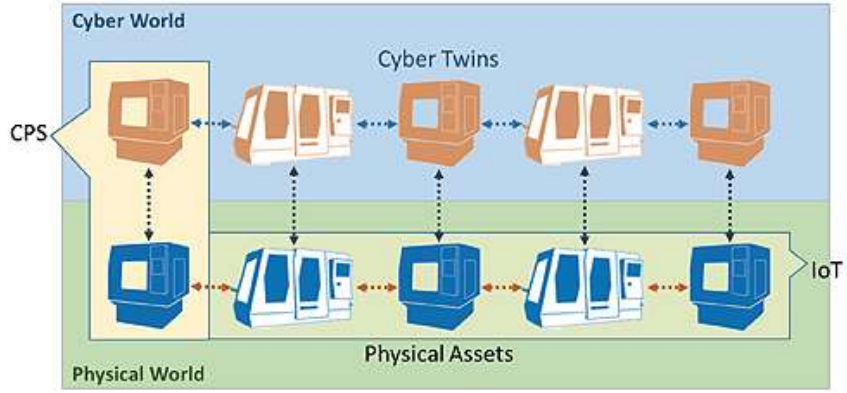
Self-adaptive systems

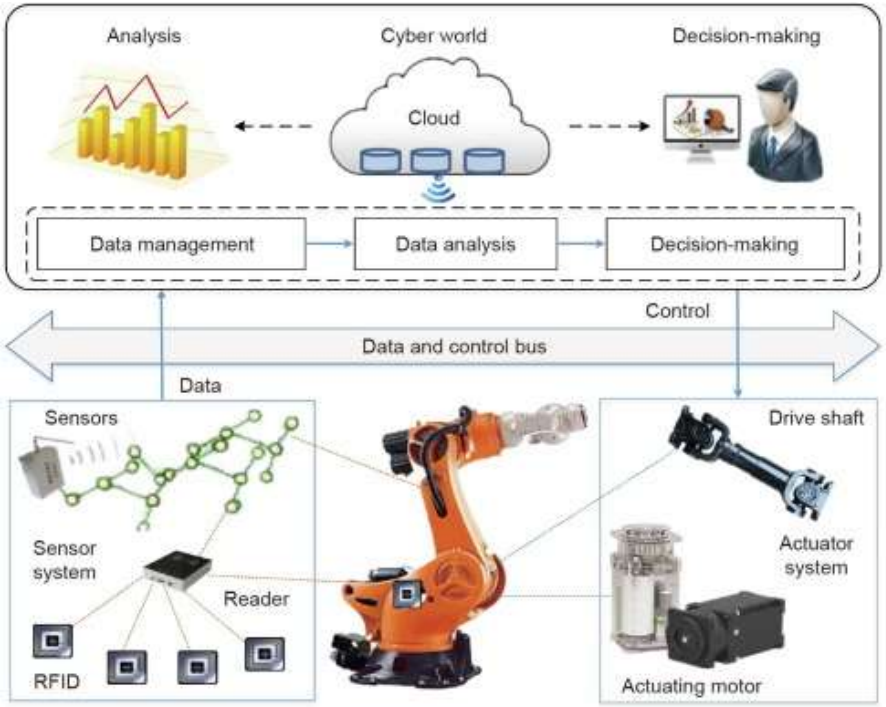


IBM Corporation. An architectural blueprint for autonomic computing (April 2003), https://www-03.ibm.com/autonomic/pdfs/AC%20Blueprint%20White%20Paper%20V7.pdf

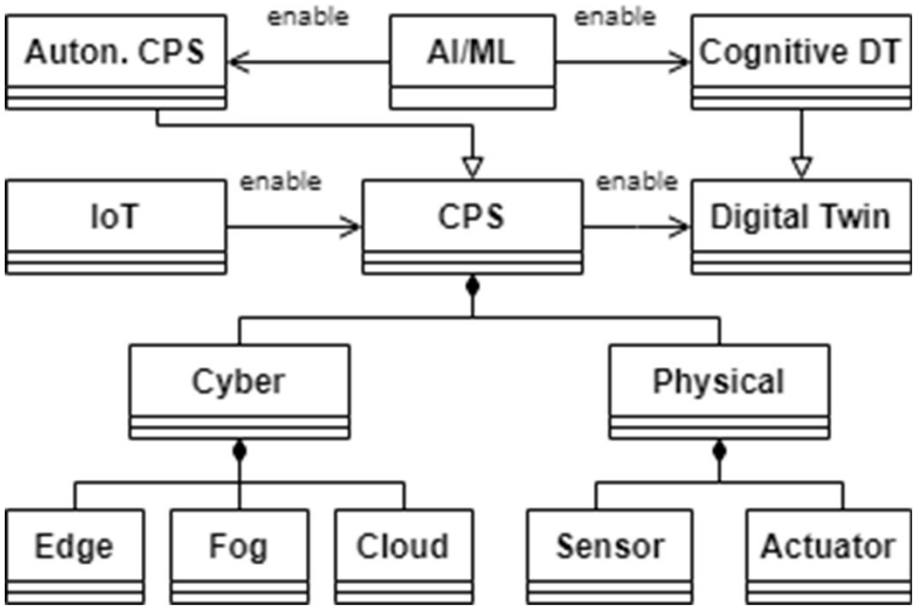
Digital Twins





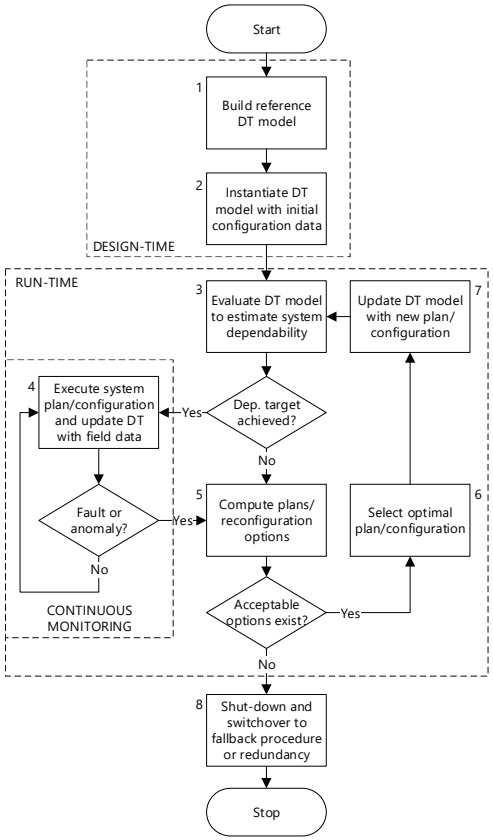


Relation among CPS, IoT, DT, and AI/ML

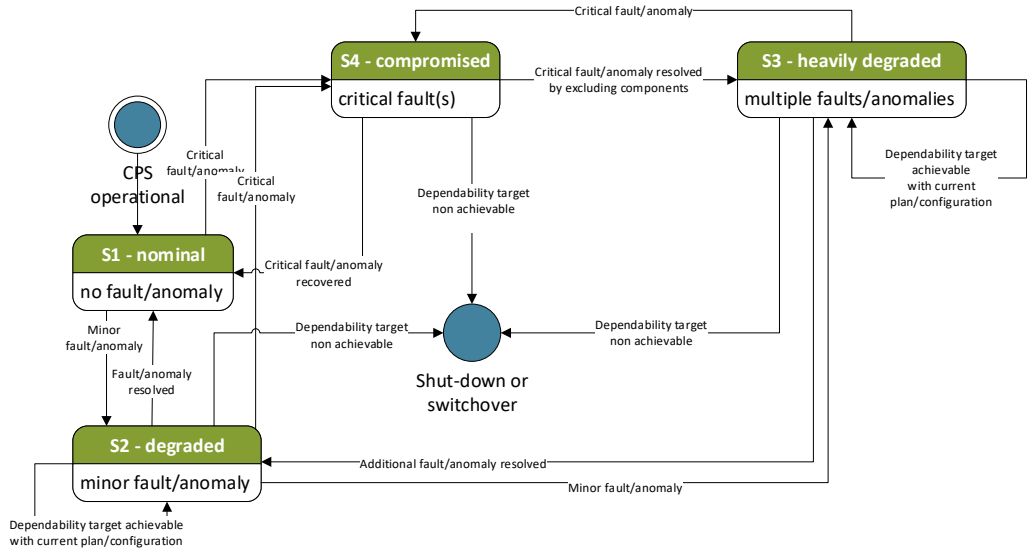


F. Flammini, C. Alcaraz, E. Bellini, S. Marrone, J. Lopez and A. Bondavalli, "Towards Trustworthy Autonomous Systems: Taxonomies and Future Perspectives," in *IEEE Transactions on Emerging Topics in Computing*, doi: 10.1109/TETC.2022.3227113.

Digital Twins as Run-Time Predictive Models

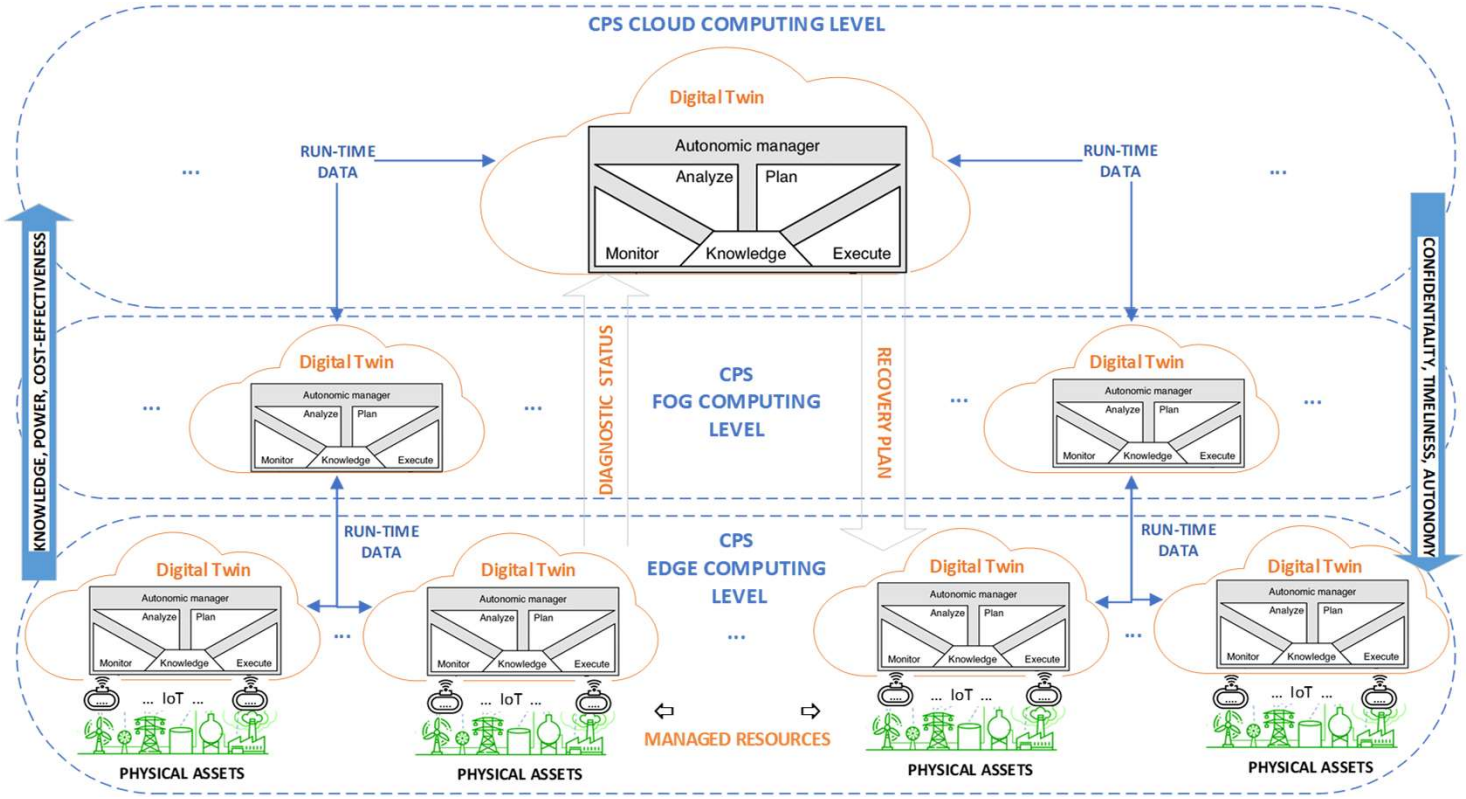


Continuous monitoring and planning and reconfiguration through DT run-time models.



State-chart describing the transitions among nominal, degraded and compromised states in self-healing CPS.

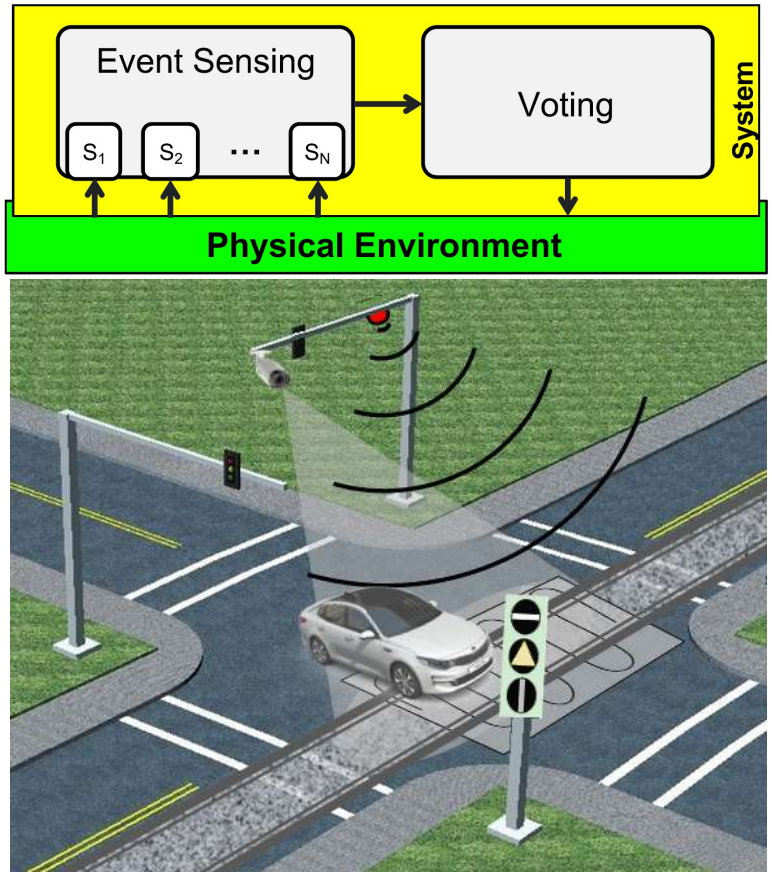
Towards a hierarchy of Digital Twins

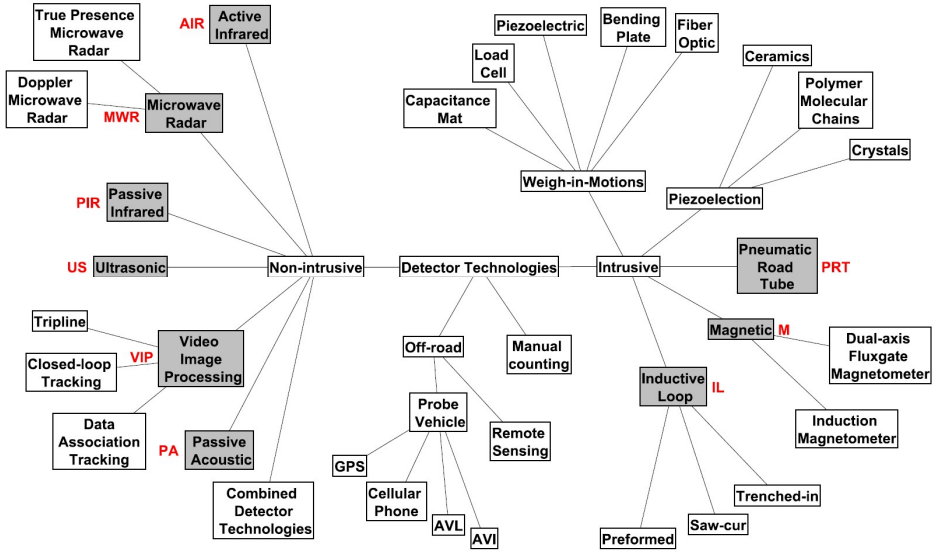


Conceptual architecture integrating the MAPE-K self-healing loop into Digital Twins at multiple levels of CPS.

Flammini, F. (2021). Digital twins as run-time predictive models for the resilience of cyber-physical systems: a conceptual framework. In: Phil. Trans. R. Soc. A. 379 http://doi.org/10.1098/rsta.2020.0369

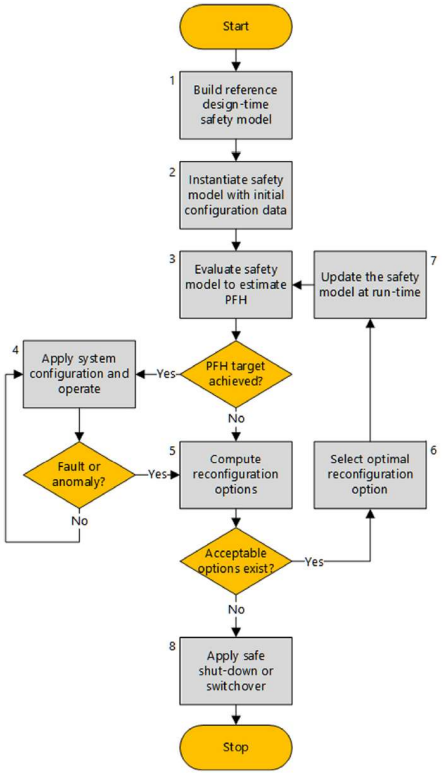
A real-world example

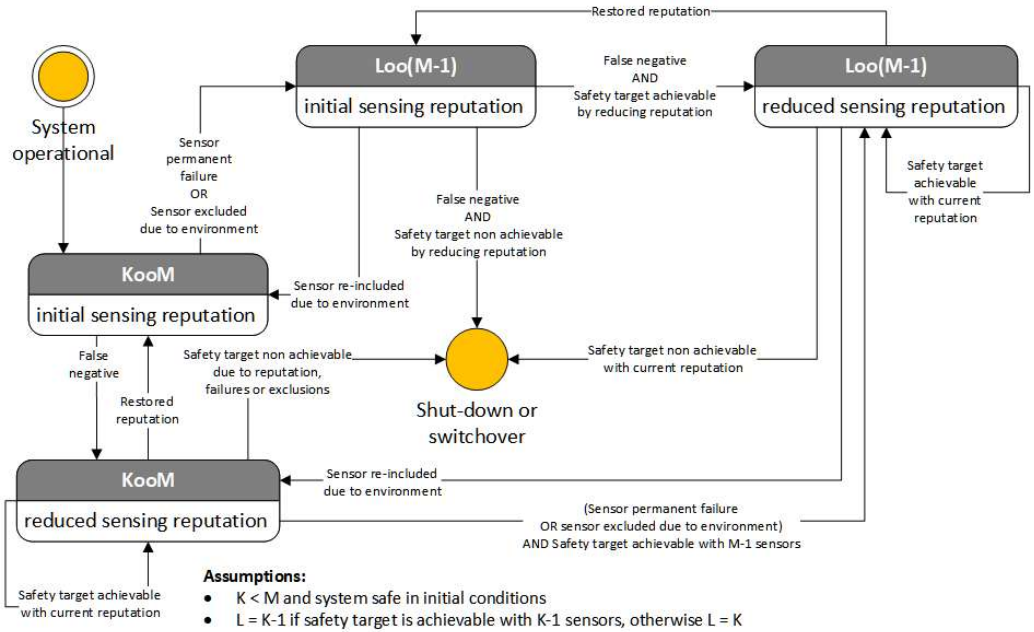




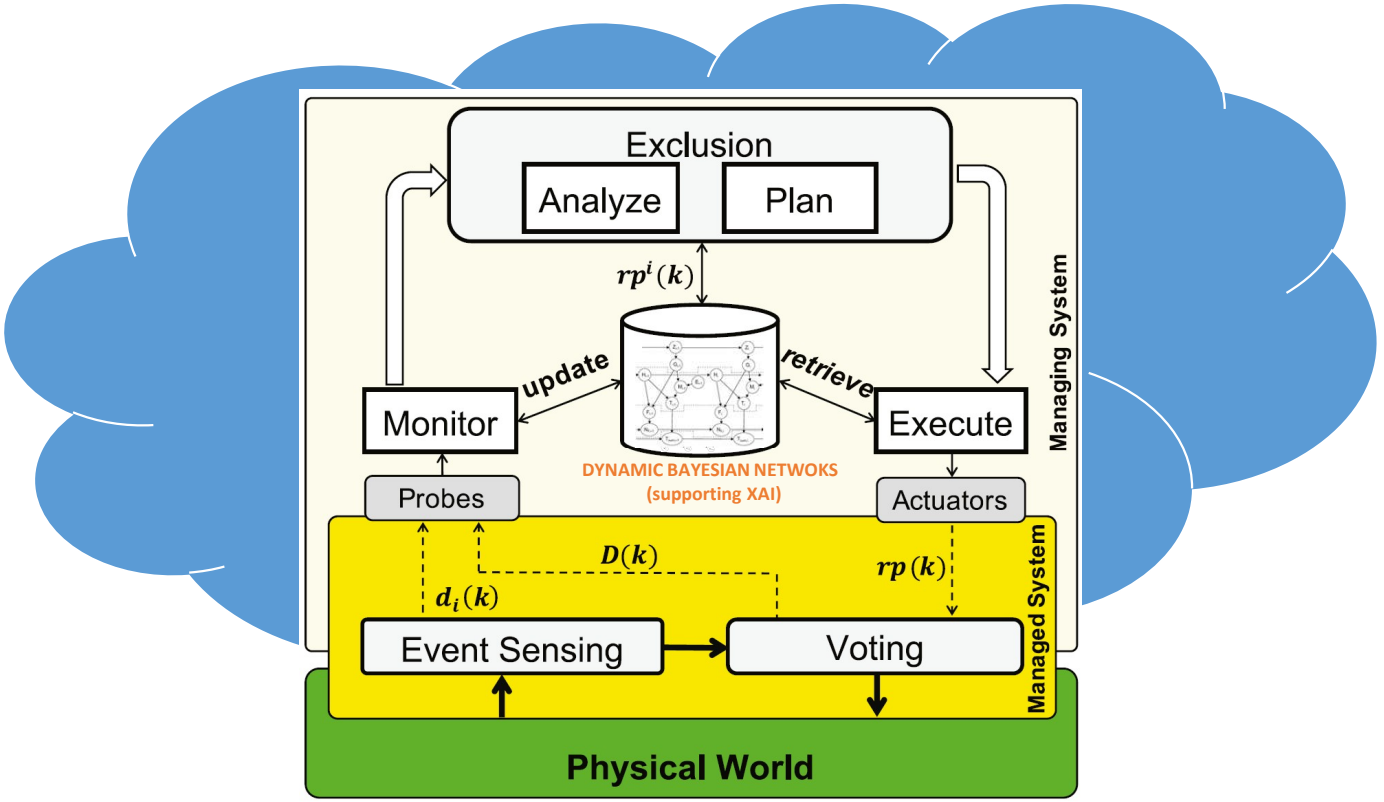
Francesco Flammini, Stefano Marrone, Roberto Nardone, Mauro Caporuscio, Mirko D'Angelo, Safety integrity through self-adaptation for multi-sensor event detection: Methodology and case-study, Future Generation Computer Systems, Volume 112, 2020, Pages 965-981, ISSN 0167-739X, https://doi.org/10.1016/j.future.2020.06.036.

Safe dynamic reconfiguration



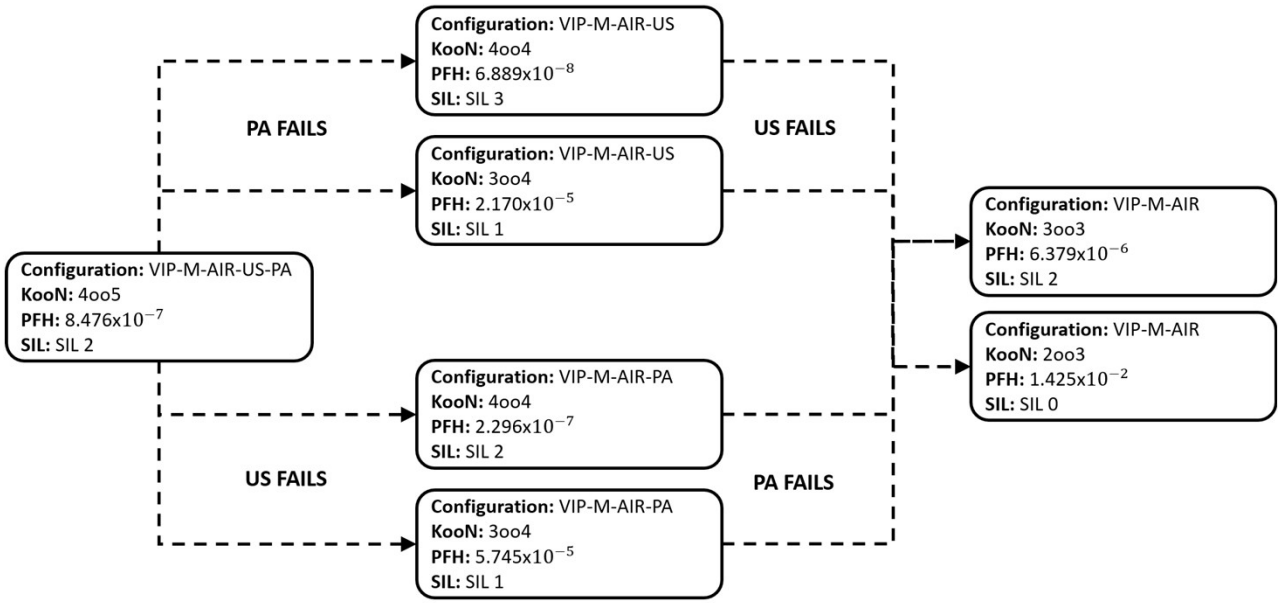


Custom MAPE-K for multi-sensor event detection



Exclusion mechanism: at every step k, the raw data di(k) captured by each sensor i, and the decision D(k) taken from Voting are used to analyze the system and compute the new sensor reputation rpi(k).

Safety integrity levels depending on reconfiguration

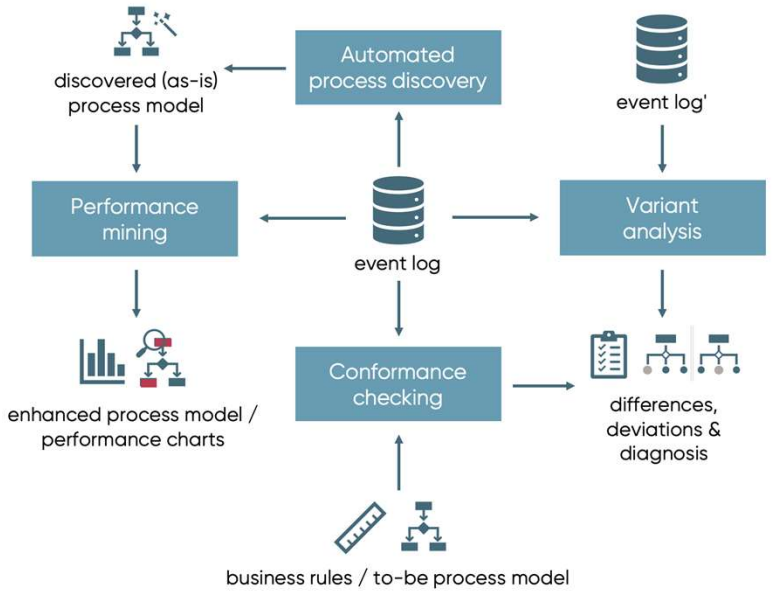


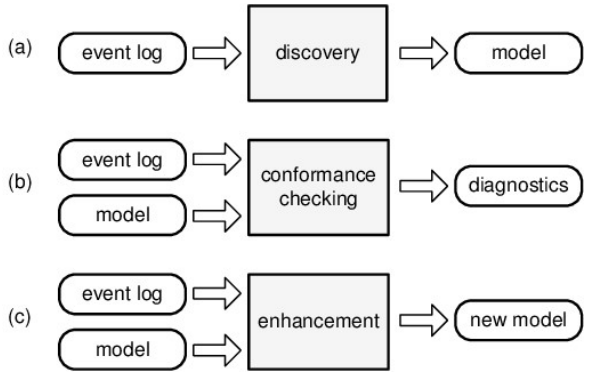
Francesco Flammini, Stefano Marrone, Roberto Nardone, Mauro Caporuscio, Mirko D'Angelo, Safety integrity through self-adaptation for multi-sensor event detection: Methodology and case-study, Future Generation Computer Systems, Volume 112, 2020, Pages 965-981, ISSN 0167-739X, https://doi.org/10.1016/j.future.2020.06.036.

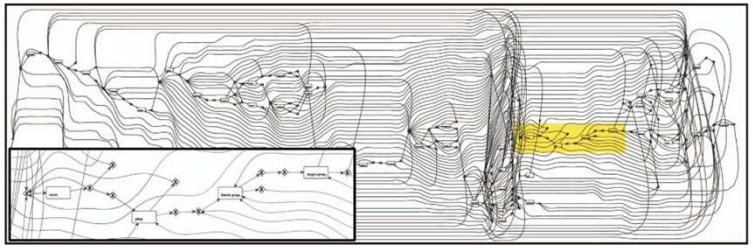
SIL	Probability of Failure per Hour (PFH)	PFH (power)	Risk Reduction Factor (RRF)
1	0.00001–0.000001	10^{-5}–10^{-6}	100,000–1, 000, 000
2	0.000001–0.0000001	10^{-6}–10^{-7}	1, 000, 000–10, 000, 000
3	0.0000001–0.00000001	10^{-7}–10^{-8}	10, 000, 000–100, 000, 000
4	0.00000001–0.000000001	10^{-8}–10^{-9}	100, 000, 000–1, 000, 000, 000

SIL requirements for continuous operation in IEC 61508

Process mining



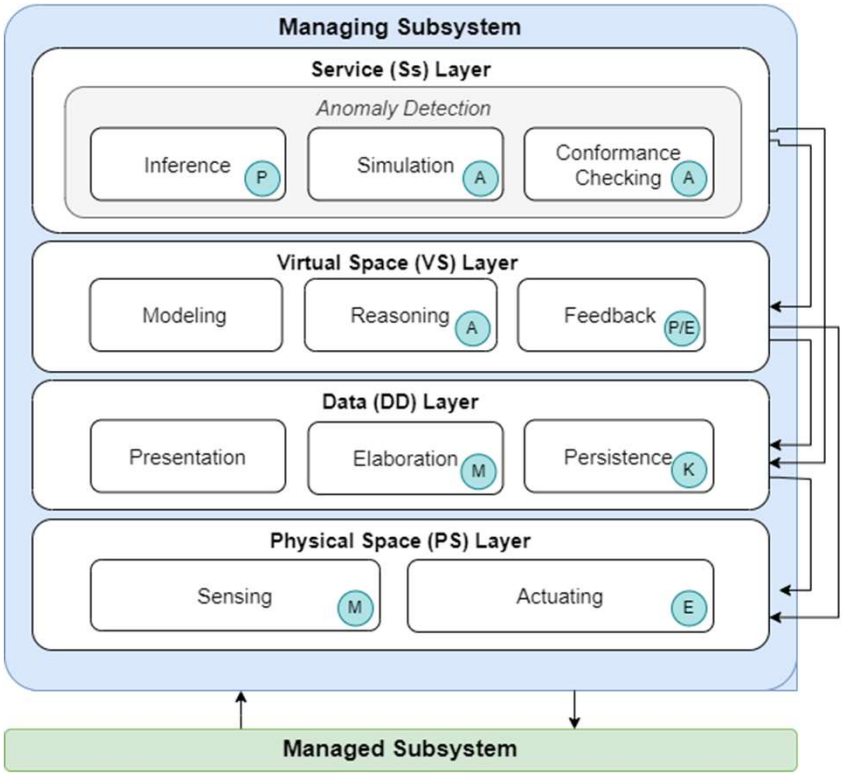




Caporuscio M, Flammini F, Khakpour N, Singh P, Thornadtsson J (2019). Smart-Troubleshooting Connected Devices: Concept, Challenges and Opportunities. Journal of Future Generation of Computer Systems (Elsevier), vol. 111, pp. 681-697, October 2020, doi: 10.1016/j.future.2019.09.004

Singh PJ, Flammini F, Caporuscio M, Saman Azari M, Thornadtsson J (2020). Towards Self-Healing in the Internet of Things by Log Analytics and Process Mining. In Proc. ESREL2020 - PSAM15, 30th European Safety and Reliability Conference and 15th Probabilistic Safety Assessment and Management Conference, Venice, Italy, June 21-26 2020

DT architecture for anomaly detection

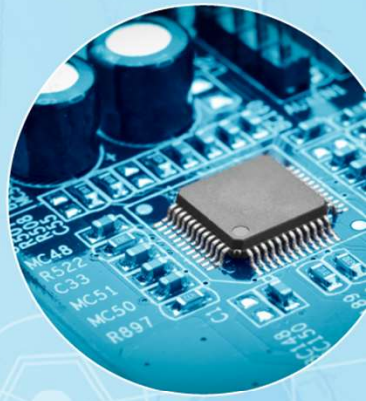


A. D. Benedictis, F. Flammini, N. Mazzocca, A. Somma and F. Vitale, "Digital Twins for Anomaly Detection in the Industrial Internet of Things: Conceptual Architecture and Proof-of-Concept," in *IEEE Transactions on Industrial Informatics*, doi: 10.1109/TII.2023.3246983.

DT-in-the-loop for mixed-reality testing of ML systems

"There is nothing permanent except change"

Heraclitus



Thank you for your kind attention!

Questions?

IEEE