Human - AI Interaction

Reflecting on freedom to reason about responsibility

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

I.R.C.C.S. Ospedale Galeazzi - Sant'Ambrogio
Gruppo San Donato

MUDI
MODELLING UNCERTAINTY, DECISIONS, INTERACTIONS

Modeling Uncertainty, Decisions and Interaction Laboratory

Artificial Intelligence and Intelligent Systems
cini National Lab

Informatica & Società
LABORATORIO NAZIONALE cini

{ReD OPEN
responsibility & design in open ecosystems

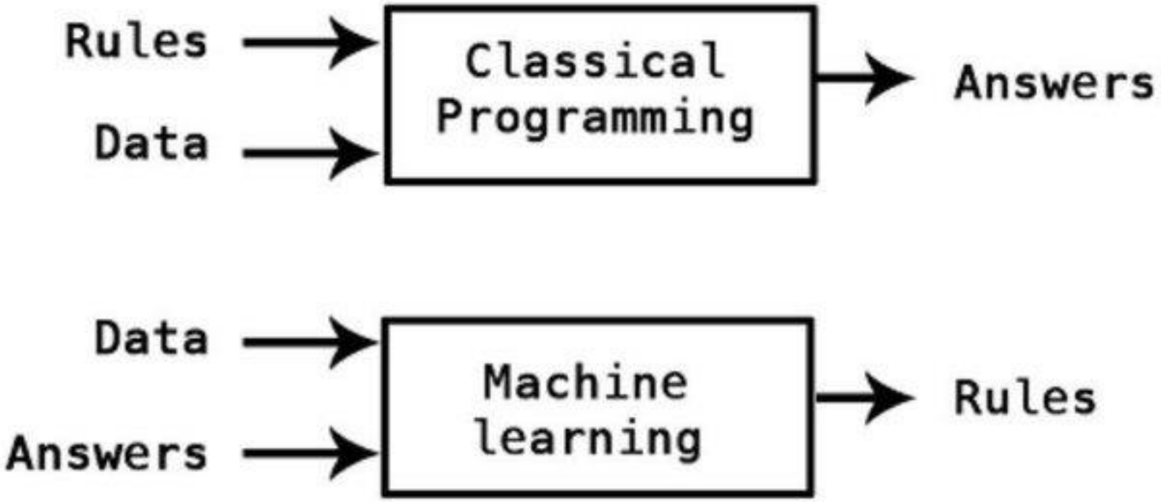
Prof. Ing. Federico Cabitza, PhD

Dipartimento di Informatica, Sistemistica e Comunicazione

Università degli Studi di Milano-Bicocca

federico.cabitza @ unimib.it

@cabitzaf



Rules

Data

Classical Programming

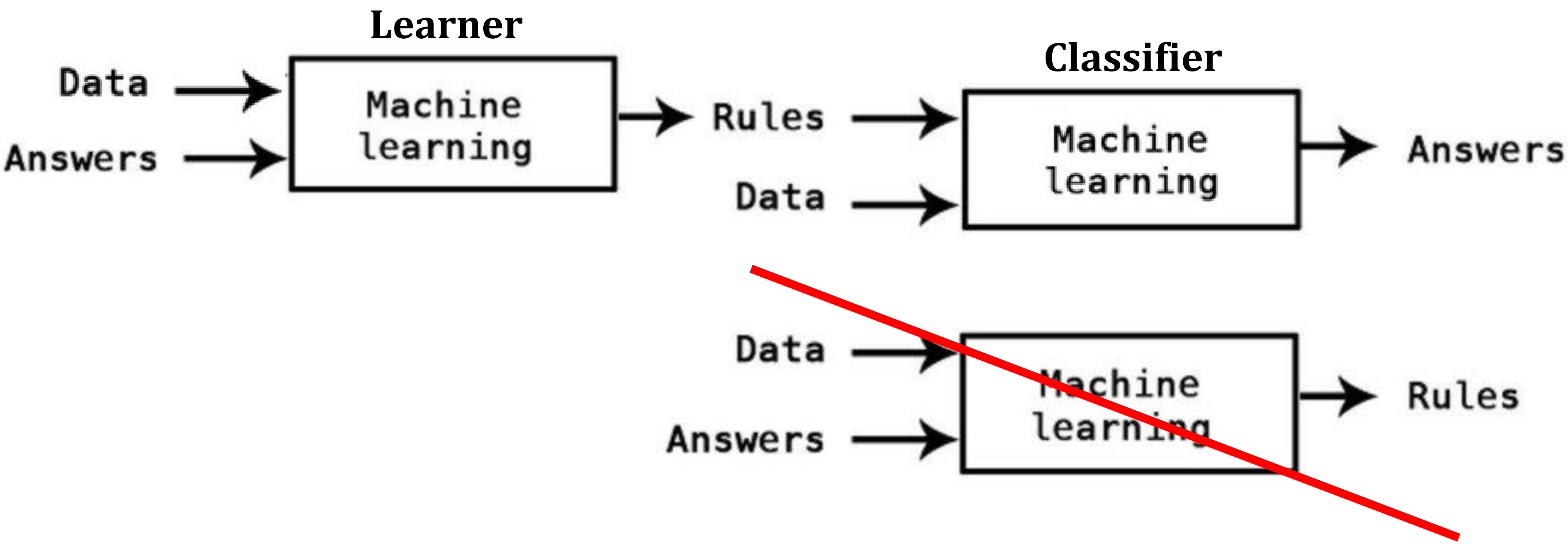
Answers

Data

Answers

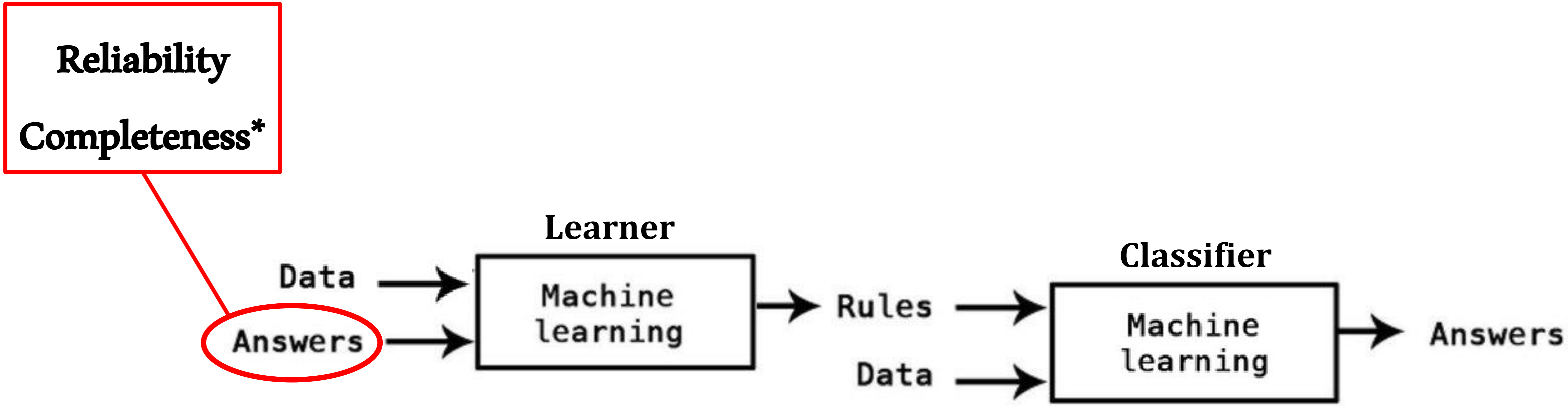
Machine learning

Rules



Learner

Classifier



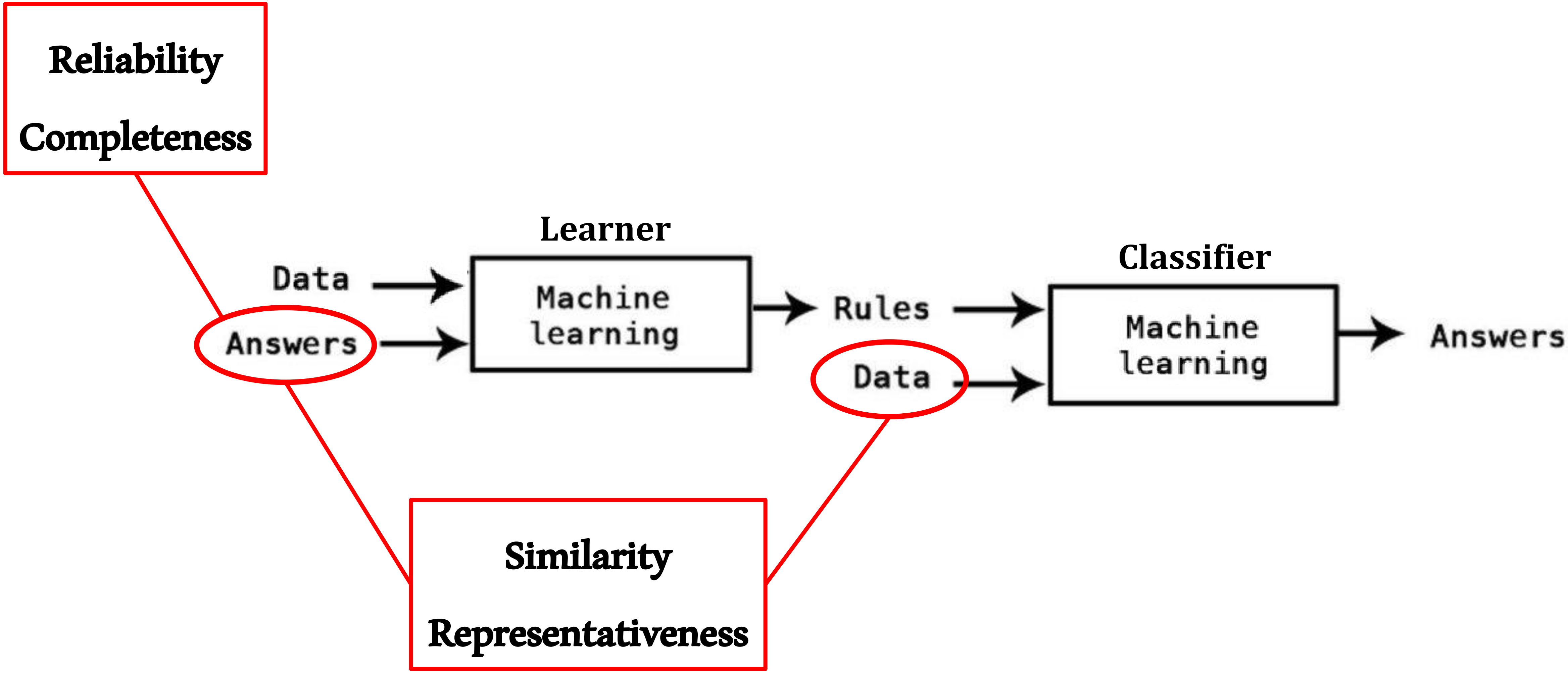
Reliability

Completeness*

Learner

Classifier

*: data minimisation or its double. “adequate, relevant and limited to what is necessary”. They must be representative (various) and enough.



Reliability

Completeness

Learner

Data

Answers

Machine learning

Rules

Classifier

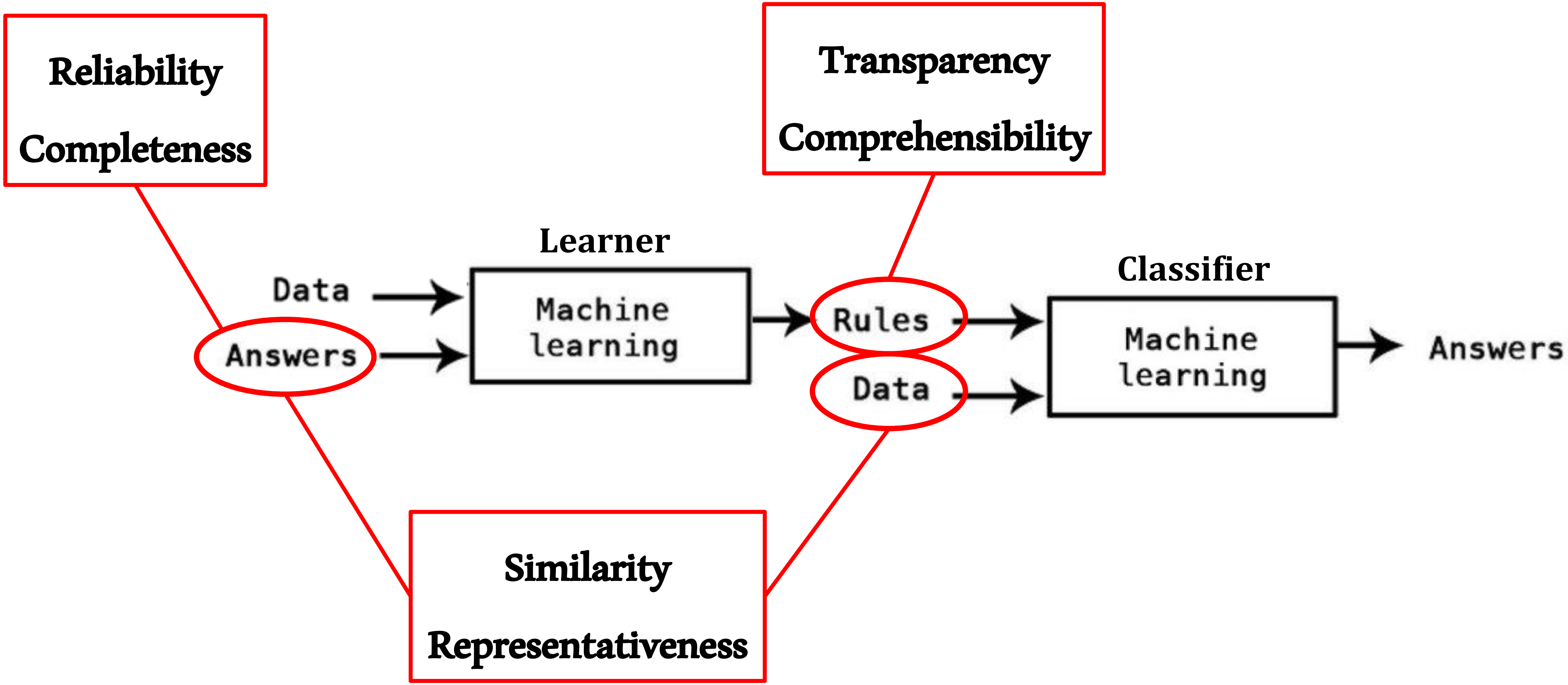
Data

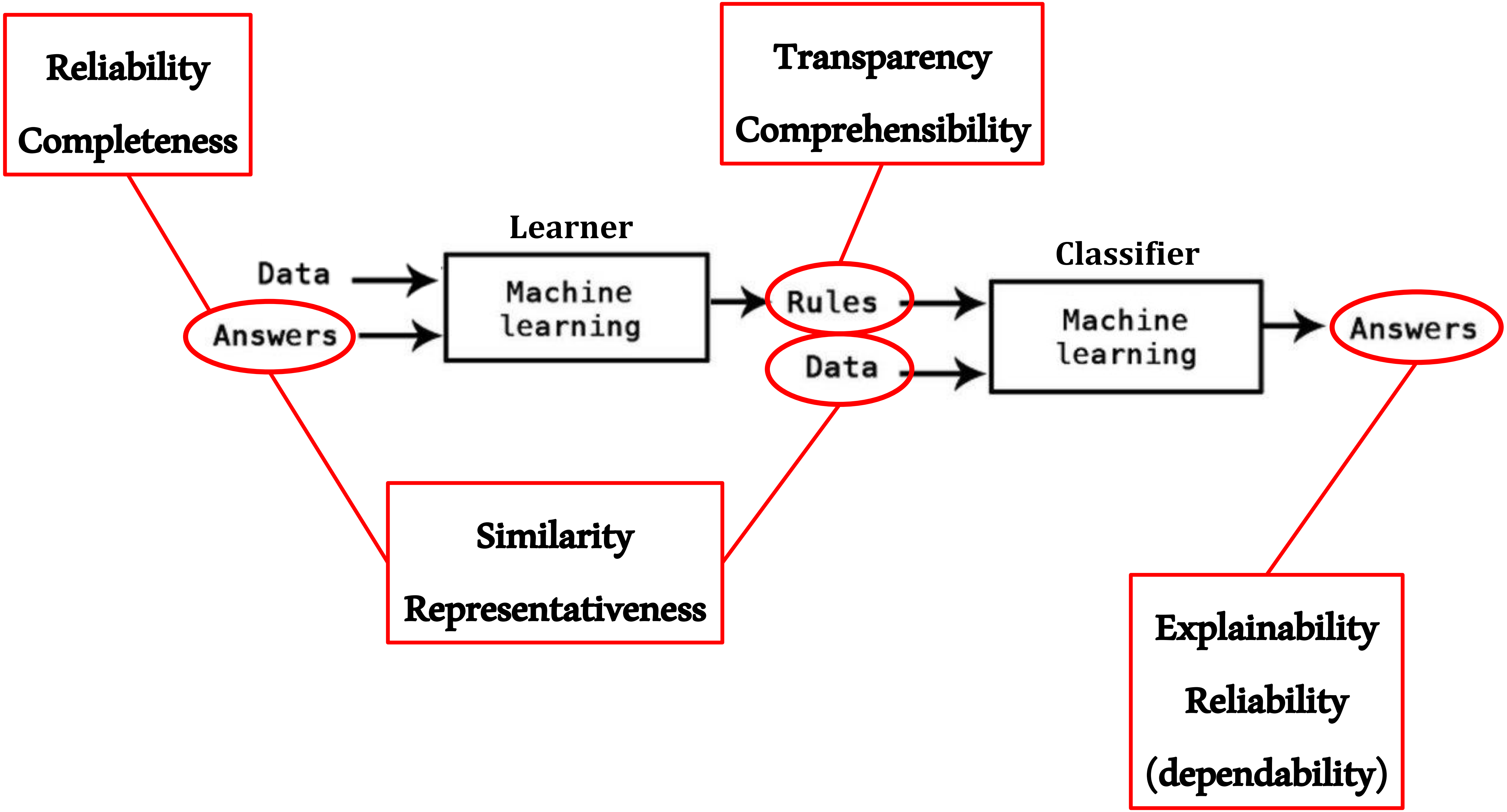
Machine learning

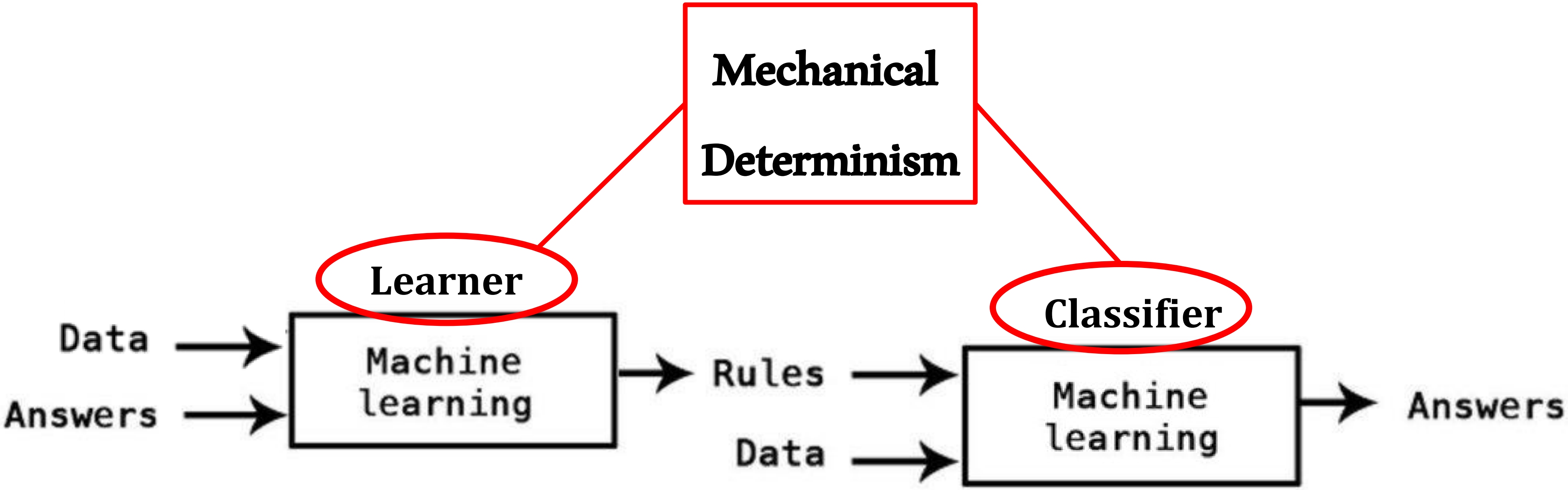
Answers

Similarity

Representativeness


Reliability
Completeness
Transparency
Comprehensibility
Learner
Data
Answers
Machine learning
Rules
Data
Classifier
Machine learning
Answers
Similarity
Representativeness


Reliability
Completeness
Transparency
Comprehensibility
Learner
Data
Answers
Machine learning
Rules
Data
Classifier
Machine learning
Answers
Similarity
Representativeness
Explainability
Reliability
(dependability)


Mechanical
Determinism
Learner
Classifier
Data
Answers
Machine
learning
Rules
Data
Machine
learning
Answers

NOTE In jurisprudence, autonomy refers to the capacity for self-governance. In this sense, also, “autonomous” is a misnomer as applied to automated AI systems, because even the most advanced AI systems are not self-governing. Rather, AI systems operate based on algorithms and otherwise obey the commands of operators. For these reasons, this document does not use the popular term autonomous to describe automation[30].

INTERNATIONAL STANDARD

ISO/IEC FDIS 22989

Information technology — Artificial intelligence — Artificial intelligence concepts and terminology

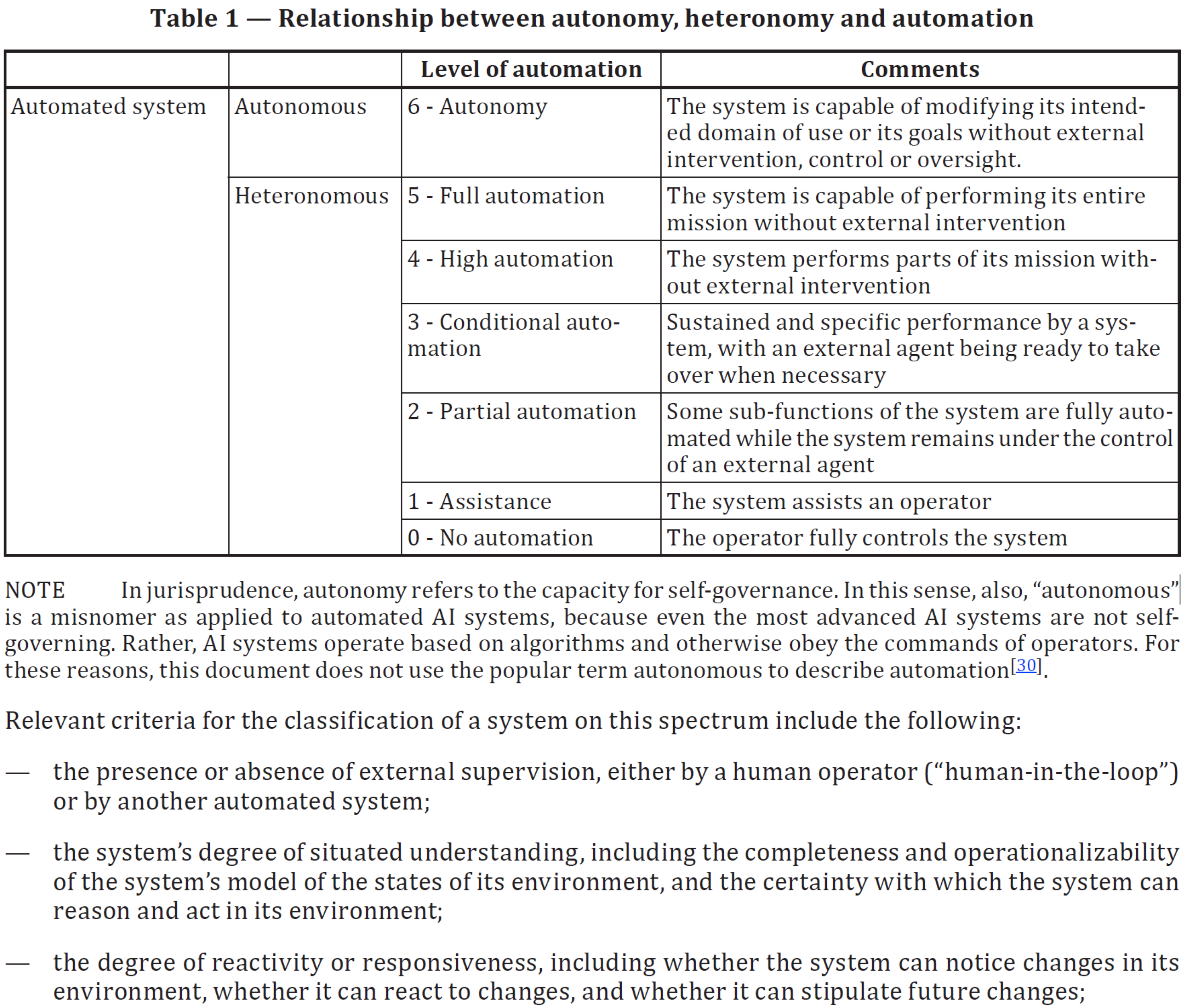


Table 1 — Relationship between autonomy, heteronomy and automation

		Level of automation	Comments
Automated system	Autonomous	6 - Autonomy	The system is capable of modifying its intended domain of use or its goals without external intervention, control or oversight.
	Heteronomous	5 - Full automation	The system is capable of performing its entire mission without external intervention
		4 - High automation	The system performs parts of its mission without external intervention
		3 - Conditional automation	Sustained and specific performance by a system, with an external agent being ready to take over when necessary
		2 - Partial automation	Some sub-functions of the system are fully automated while the system remains under the control of an external agent
		1 - Assistance	The system assists an operator
		0 - No automation	The operator fully controls the system

NOTE In jurisprudence, autonomy refers to the capacity for self-governance. In this sense, also, “autonomous” is a misnomer as applied to automated AI systems, because even the most advanced AI systems are not self-governing. Rather, AI systems operate based on algorithms and otherwise obey the commands of operators. For these reasons, this document does not use the popular term autonomous to describe automation[30].

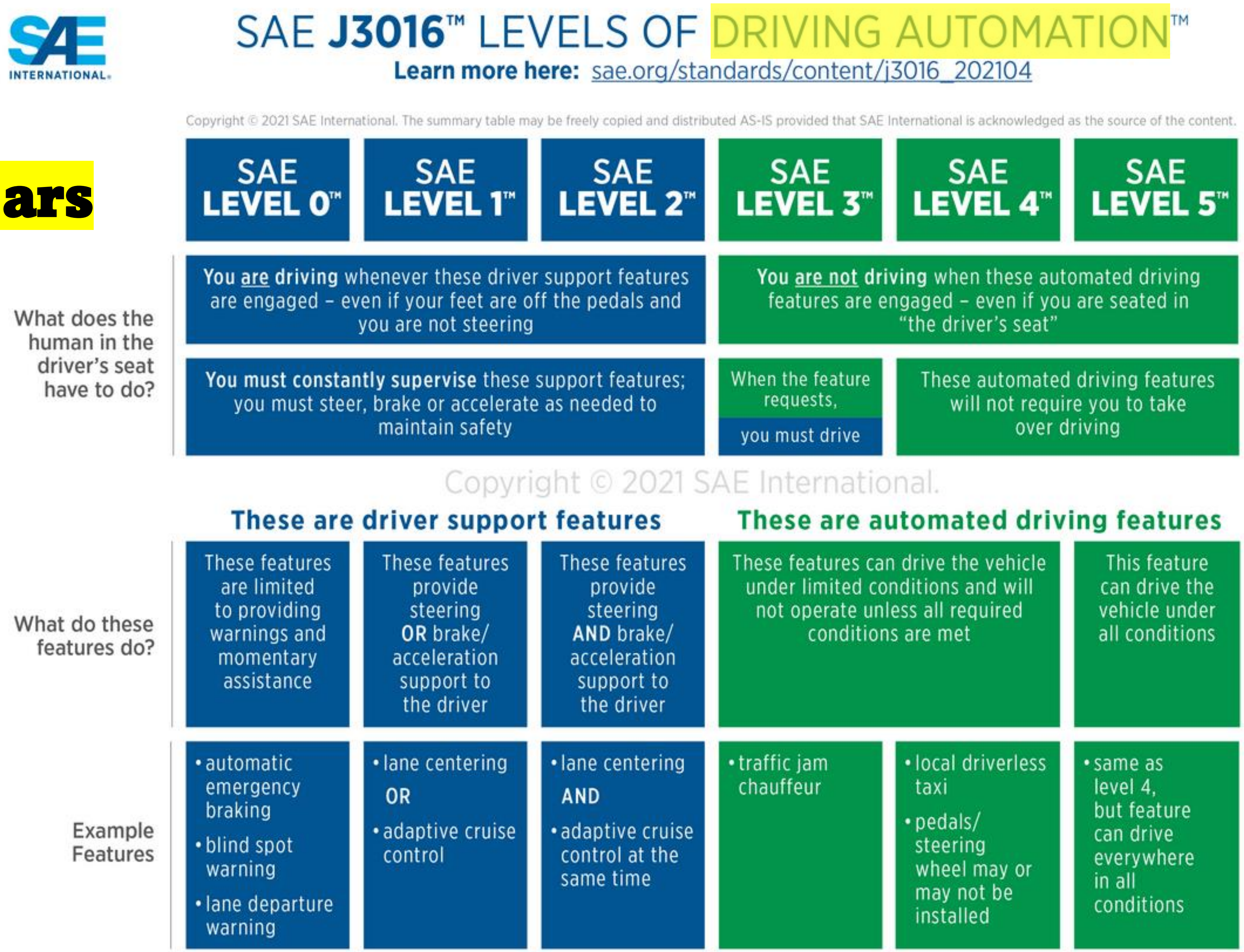
Relevant criteria for the classification of a system on this spectrum include the following:

- the presence or absence of external supervision, either by a human operator (“human-in-the-loop”) or by another automated system;
- the system’s degree of situated understanding, including the completeness and operationalizability of the system’s model of the states of its environment, and the certainty with which the system can reason and act in its environment;
- the degree of reactivity or responsiveness, including whether the system can notice changes in its environment, whether it can react to changes, and whether it can stipulate future changes;

Autonomous Cars

Autonomous Cars

Regulation (EU) 2019/2144
"automated vehicle", "fully automated vehicle": "designed and constructed to move autonomously without any driver supervision"



SAE INTERNATIONAL

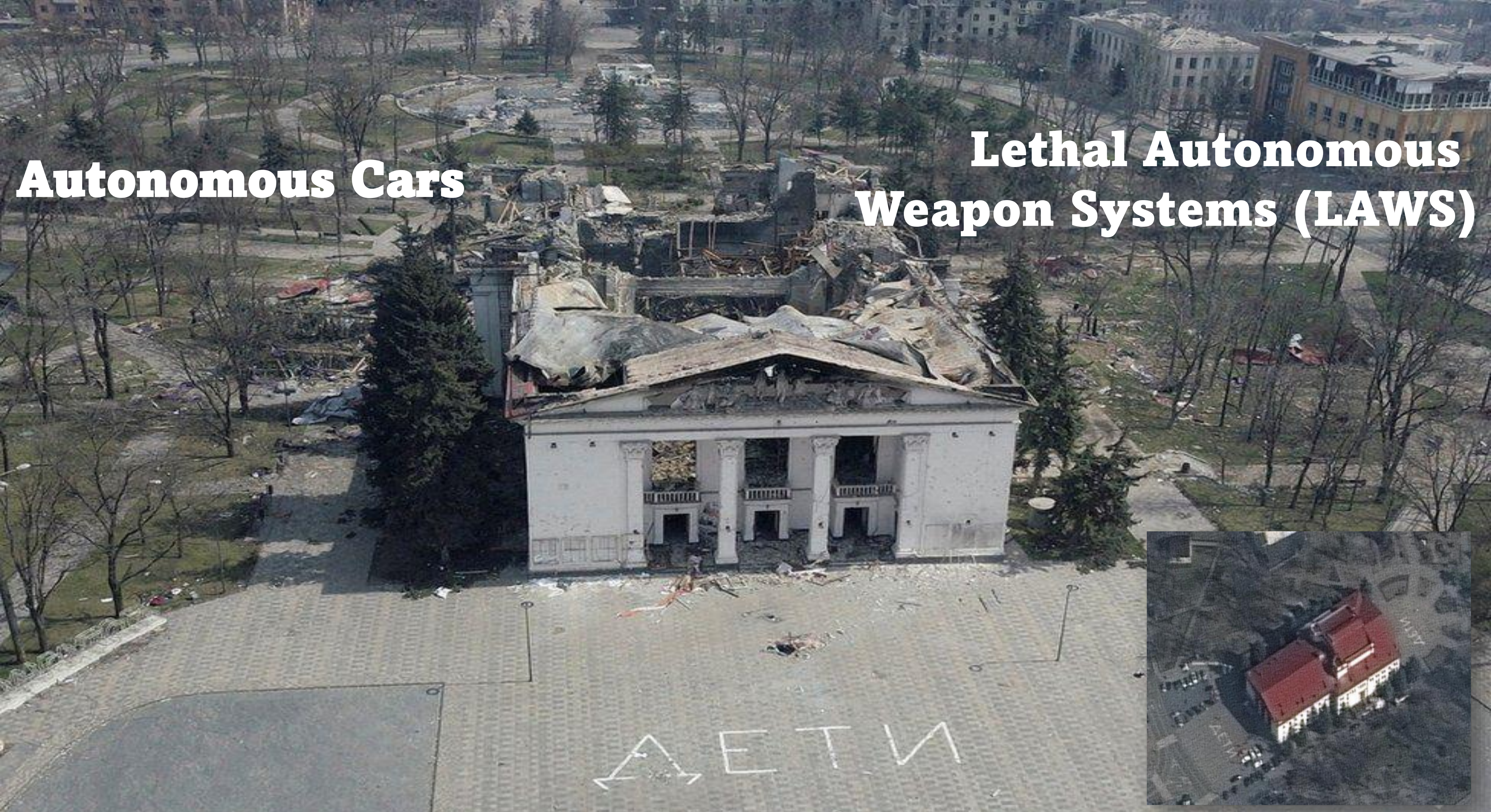
SAE **J3016**™ LEVELS OF DRIVING AUTOMATION™

Learn more here: sae.org/standards/content/j3016_202104

	SAE **LEVEL 0**™	SAE **LEVEL 1**™	SAE **LEVEL 2**™	SAE **LEVEL 3**™	SAE **LEVEL 4**™	SAE **LEVEL 5**™
What does the human in the driver's seat have to do?	**You are driving** whenever these driver support features are engaged – even if your feet are off the pedals and you are not steering			**You are not driving** when these automated driving features are engaged – even if you are seated in "the driver's seat"		
	You must constantly supervise these support features; you must steer, brake or accelerate as needed to maintain safety			When the feature requests, you must drive	These automated driving features will not require you to take over driving	
	These are driver support features			**These are automated driving features**		
What do these features do?	These features are limited to providing warnings and momentary assistance	These features provide steering **OR** brake/ acceleration support to the driver	These features provide steering **AND** brake/ acceleration support to the driver	These features can drive the vehicle under limited conditions and will not operate unless all required conditions are met		This feature can drive the vehicle under all conditions
Example Features	• automatic emergency braking • blind spot warning • lane departure warning	• lane centering **OR** • adaptive cruise control	• lane centering **AND** • adaptive cruise control at the same time	• traffic jam chauffeur	• local driverless taxi • pedals/ steering wheel may or may not be installed	• same as level 4, but feature can drive everywhere in all conditions

Autonomous Cars

Lethal Autonomous Weapon Systems (LAWS)



Mariupol theatre; children

Autonomous Cars

Lethal Autonomous Weapon Systems (LAWS)

Can these machines decide to run over a pedestrian or spare a civilian target ?

I'm sorry Dave, I'm afraid I can't do that

αὐτονομία

αὐτονομία

τά νόμιμα

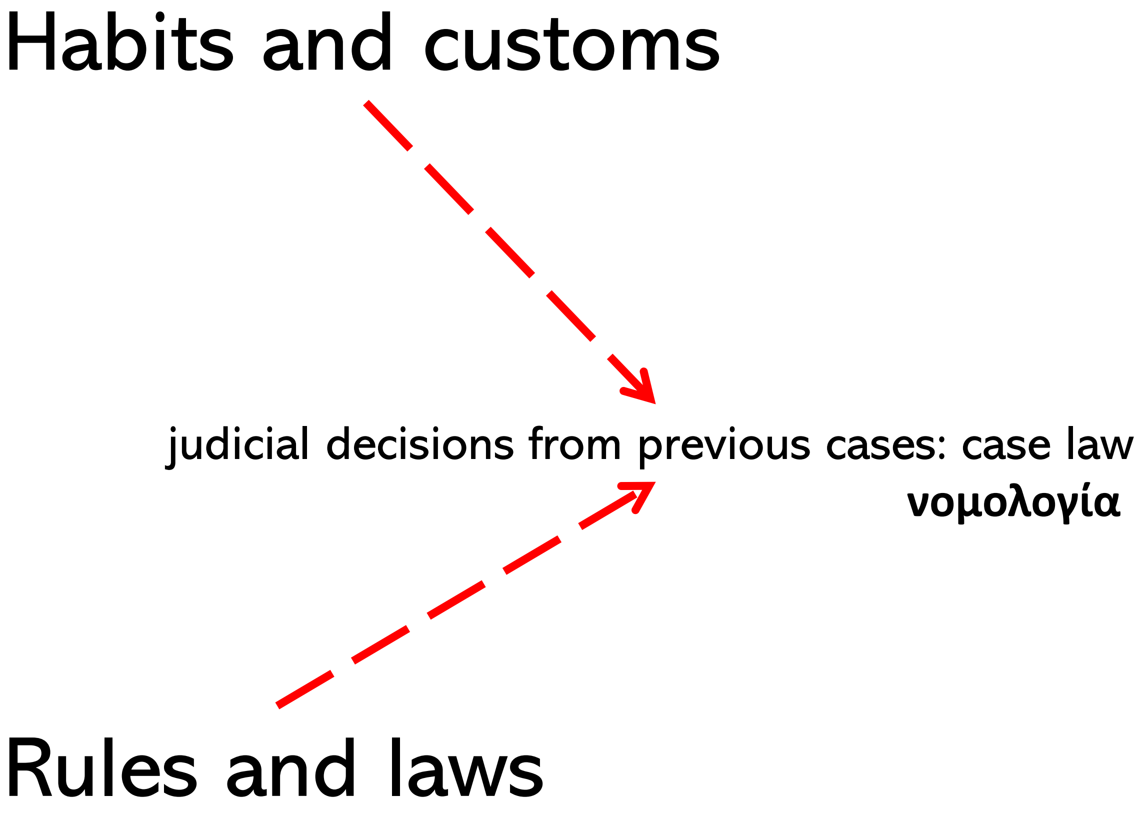
Habits and customs

αὐτονομία

Rules and laws

Νέμω, I distribute
Θεσμοί,
τίθημι, I set out, assign

αὐτονομία



Habits and customs

judicial decisions from previous cases: case law

νομολογία

Rules and laws

αὐτονομία

αὐτονομία

Who? What?

αὐτονομία

A material object, an artifact, a digital device? Or the connected socio-technical system? (technology as always technology-in-use within a community of competent actors)

αὐτονομία

A material object, an artifact, a digital de

c

te

(t

al

us

community of competent actors)

An actor (actant) an entity that acts, and in so doing, it modifies another entity. It does not pre-exist this relation of influence, withou the network (rhizome?) binding it to other nodes. Even more, the actor, not as a stable, firm entity, but as a more-or-less temporary assemblage, as a «stream».

αὐτονομία

A material object, an
artif
de
co
te
(t
al
us
community of
competent actors)

Technology as «instrumentation of human action» [1] or even as “human behavior” [2] that transforms society and the environment.
Structured/ing behavior that exerts agency.

[1] Johnson, Deborah (1985). Computer ethics. *Englewood Cliffs (NJ)*, *10*, 102926.

[2] Devon, Richard and Van de Poel Ibo (2004) Design Ethics: The Social Ethics Paradigm. International Journal of Engineering Education

αὐτονομία

From
«humans in the loop»
To
«computers in the group»

A material object, an
artif
de
c
te
(t
al
us
community of
competent actors)

Technology as «instrumentation of human action» [1] or even as "human behavior" [2] that transforms society and the environment.
Structured/ing behavior that exerts agency.

International Journal of Human-Computer Studies

Volume 155, November 2021, 102696

The need to move away from agential-AI: Empirical investigations, useful concepts and open issues

Federico Cabitza [a], Andrea Campagner [a], Carla Simone [b]

αὐτονομία

From
«humans in the loop»
To
«computers in the group»

BA 場

References

- Cabitza, F., Campagner, A., & Simone, C. (2021). The need to move away from agential-AI: Empirical investigations, useful concepts and open issues. *International Journal of Human-Computer Studies*, *155*, 102696.
- Cabitza, F., Rasoini, R., & Gensini, G. F. (2017). Unintended consequences of machine learning in medicine. *Jama*, *318*(6), 517-518.
- Cabitza, F. (2021). Cobra AI: Exploring Some Unintended Consequences. *Machines We Trust: Perspectives on Dependable AI*, 87.
- Cabitza, F. (2022) Intelligenza Artificiale e deskilling decisionale. MIT Sloan Management Review Italia. 1(2)
- Cabitza F. et al. (2023) AI shall have no dominion: on how to measure technology dominance in AI-supported human decision making. Forthcoming.