4
Key topics

Due to its scope and complexity, it seems reasonable to structure the topic of AI according to basic topics, horizontal topics, as well as relevant economic and application areas (see Figure 6).

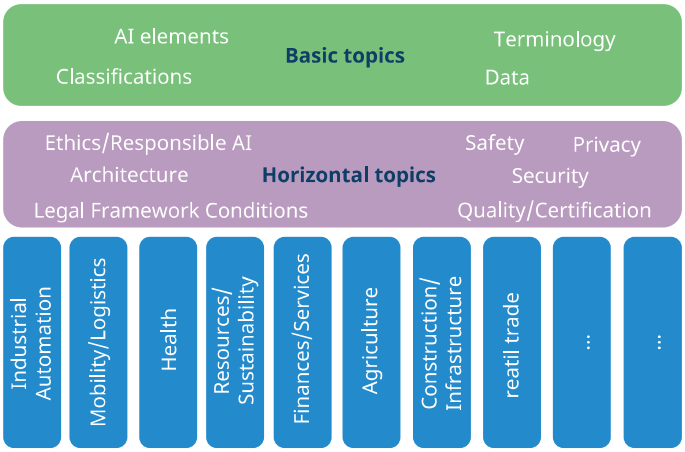


Figure 6: Chart of basic topics, horizontal topics, and relevant economic and application areas

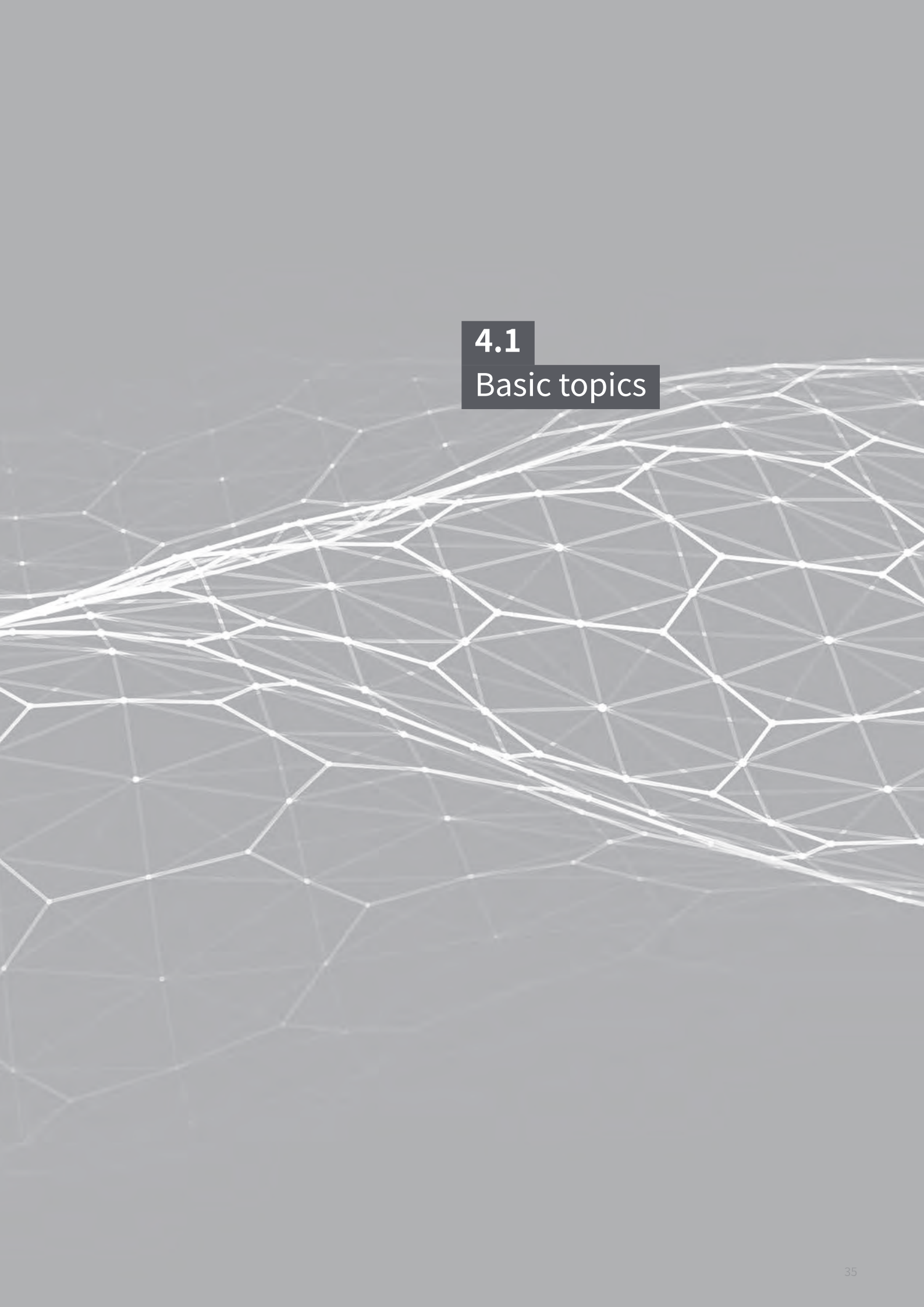
The basic topics form the basis for discussions on AI. This includes, for example, terminology (definitions), classifications, but also topics such as data (data analyses, data formats, data quality, etc.).

New technical developments, especially in the application of AI, raise new questions on overarching issues such as IT security, quality, ethics or the legal framework. Ethical aspects of responsibility in the use of AI technologies, as well as issues such as fairness, security, social inclusion and transparency of algorithms must be considered. In addition, the foundations for cross-industry quality criteria must be developed to enable the analysis and certification of AI systems. Which legal relationship AI may have in the future is another cross-sectional topic to be discussed.

The economic fields of application for AI are extremely diverse. AI is relevant for almost all sectors of the economy, and also for other areas of application outside the economy, and is found both in the form of components in end products and services, and in the productive core processes and support processes within companies.

In this first version, the present Standardization Roadmap AI focuses on the areas of basic topics, horizontal topics (ethics, quality/conformity assessment/certification, IT security) and the three application areas of industrial automation, mobility/logistics and health.

In the following Chapter 4.1 to 4.7, the starting situation, challenges and essential standardization needs of the seven main topics are elaborated.



4.1 Basic topics

Definition of the term "artificial intelligence"

Providing a precise definition of the term "artificial intelligence" is a difficult task due to a multitude of different perspectives and opinions on this topic:

1. Does the term refer to a scientific or technical discipline, or does it refer to a system property or capability?
2. Should the term be limited to a description of the function of AI systems or refer to their implementation?
3. Should terms commonly associated with human intelligence (like "knowledge", "skills") be used to explain AI?

Almost every organization dealing with AI defines this term in different ways. In view of the difficulties in finding a generally accepted definition, this will not be done in this document. 4.1.2.1 gives an overview of the different classes of AI methods and their capabilities and areas of application, which will be used for the following discussion to narrow down the term. However, the range of possible definitions of AI is illustrated by the following examples in Table 1:

Table 1: Different definitions of AI

Example	German	English	Source
1	Künstliche Intelligenz beschreibt Systeme, die intelligentes Verhalten dadurch zeigen, dass sie – mit einem gewissen Grad an Autonomie – ihre Umgebung analysieren und entsprechend agieren, um spezifische Ziele zu erreichen.	Artificial intelligence (AI) refers to systems that display intelligent behaviour by analyzing their environment and taking actions – with some degree of autonomy – to achieve specific goals.	[30]
2	<System> Fähigkeit, sich Wissen anzueignen, zu verarbeiten, zu kreieren und anzuwenden, das in einem Modell gespeichert wird, um eine oder mehrere vorgegebene Aufgaben zu erfüllen <Technische Disziplin> Disziplin zur Entwicklung und Erforschung von KI Systemen <Künstliche Intelligenz> Informationen zu Objekten, Ereignissen, Konzepten oder Regeln, ihren Beziehungen und Eigenschaften, zur zielorientierten Nutzung organisiert Anmerkung 1 zum Begriff: Information kann in numerischer oder symbolischer Form existieren. Anmerkung 2 zum Begriff: Informationen sind kontextualisierte Daten, die damit interpretierbar werden. Daten werden durch Abstraktion oder durch Messungen der Umgebung kreiert.	<system> capability to acquire, process, create and apply knowledge, held in the form of a model, to conduct one or more given tasks <engineering discipline> discipline of developing and studying AI systems <artificial intelligence> information about objects, events, concepts or rules, their relationships and properties, organized for goal-oriented systematic use Note 1 to entry: Information may exist in numeric or symbolic form. Note 2 to entry: Information is data that has been contextualized, so that it is interpretable. Data are created through abstraction or measurement from the world.	ISO/CD 22989, ongoing project in ISO/IEC JTC 1/SC 42, currently at Committee Draft (CD) stage
3	Das Design und die Konstruktion intelligenter Agenten, die Wahrnehmungen ihrer Umgebung erhalten und deren Handlungen ihre Umgebung beeinflussen.	The designing and building of intelligent agents that receive percepts from the environment and take actions that affect that environment.	[31]
4	Ein KI-System ist ein maschinenbasiertes System, das in der Lage ist, für eine vorgegebene Menge von durch den Menschen definierte Ziele Vorhersagen, Empfehlungen oder Entscheidungen, die reale oder virtuelle Umgebungen beeinflussen, vorzunehmen. KI-Systeme werden entwickelt, um mit verschiedenen Graden von Autonomie zu operieren.	An AI system is a machine-based system that can, for a given set of human defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy.	[32]
5	Künstliche Intelligenz (KI) ist ein Teilgebiet der Informatik mit dem Ziel, intelligentes Verhalten und die zugrundeliegenden kognitiven Fähigkeiten auf digitalen Computern zu realisieren.	Artificial intelligence (AI) is a branch of computer science with the goal of realizing intelligent behaviour and the underlying cognitive abilities on digital computers.	[33]

Autonomous systems [33] can independently solve complex tasks in a specific application domain despite varying objectives and initial situations. Autonomous systems must independently generate an action plan, depending on the current task context, with which an overall goal specified by the operator of the autonomous system can be achieved without remote control and, if possible, without the intervention and assistance of human operators within the framework of legal and ethical requirements. If individual actions of the autonomous system fail during the execution of the plan, the system must be able to carry out a plan revision on its own in order to achieve the specified objective by adapting the original plan in another way. A new generation of autonomous systems is also able to solve a distributed task together with other autonomous systems and/or a group of people. Within the framework of self-regulation, an autonomous system must also have explicit models of its own performance limits and, in the case of specifications or environmental conditions that do not indicate a successful autonomous achievement of objectives, must inform the system operator of this fact (e.g. excessive shear winds prevent drone flight, an extremely steep section of the route exceeds the maximum climbing capacity of an autonomous vehicle).

A reference architecture for autonomous systems has been developed in the High-Tech Forum of the German Federal Government (see Figure 7). It is based on sensors for observing the environment and actuators for changing environmental conditions in order to achieve the objectives of the autonomous system. In addition, communication with the networked environment of the system and with cooperating humans can provide further important information for the behaviour of the autonomous system. In principle, the autonomous system consists of several modules for cognitive information processing, which are controlled by different mechanisms for self-regulation, as well as several knowledge bases, which are constantly adapted by machine learning and reasoning starting from an initial configuration.

With knowledge bases, an episodic memory serves as a long-term memory for events that have directly affected the autonomous system to enable case-based reasoning and learning from experience. The entire course of the system's communication with humans and technical systems in the environment is stored in the discourse memory, so that references to the aforementioned and ambiguities in the context can be resolved at any time. A plan library stores successfully executed plans for common classes of problems in order to achieve goals more efficiently through plan revision without replanning, and through plan recognition based on observing the actions of other agents in the environment to identify their intent.

Domain models contain networked models of all relevant objects, relations, states and events in an application field, which are necessary for their recognition by sensors or for their transformation by the actuators of the autonomous system. In task models, typical task classes for an autonomous system are schematically recorded in order to quickly understand and classify a new task set by the system operator or to decompose it into a series of known tasks. User models are particularly crucial when using autonomous systems as assistance systems in the service sector, since they contain assumptions about the preferences, abilities

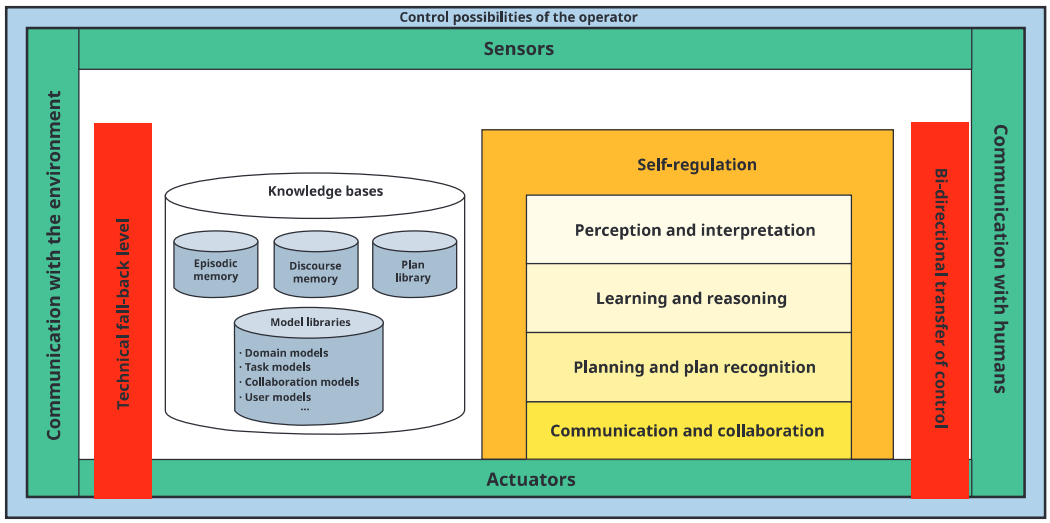


Figure 7: Reference architecture for autonomous systems [33]

and level of knowledge of a system user, among other things, which enable personalization of service performance through adaptive behaviour.

In order to increase confidence in the use of autonomous systems and to minimize the risk of endangering people in the environment in the event of a complete technical failure of the central control functions, there must be a technical fall-back level in accordance with the reference architecture which, in an emergency, puts the autonomous system into a safe operating state, for example via a redundant mechatronic function or a radio-based remote control, and generates an alarm message via communication with the environment.

More often, an autonomous system will reach the limits of its abilities in abnormal situations and will have to hand over control to a human being. A bi-directional transfer of control must be provided for, so that after overcoming an obstacle to the achievement of objectives which cannot be achieved by the autonomous system alone, the human being can completely return control to the system.

4.1.1 Status quo

With regard to AI basic topics, the SC 42 is carrying out work on various documents:

- → **ISO/IEC 22989, Artificial intelligence – Concepts and terminology.** This standard is being developed under the leadership of a German editor.
- → **ISO/IEC 23053, Framework for Artificial Intelligence (AI) Systems Using Machine Learning** (ML) describes a terminological framework for machine learning.
- → **ISO/IEC 23894, Information Technology – Artificial Intelligence – Risk Management** contains guidelines for the risk management for the development and use of AI systems. This standard, too, is being developed under the leadership of a German editor.
- → **ISO/IEC 38507, Information technology – Governance of IT – Governance implications of the use of artificial intelligence by organizations** deals with organizational governance in connection with AI.
- → **ISO/IEC 20546, Information technology – Big data – Overview and vocabulary** [34] deals with concepts and terminology relating to big data, which is also being considered within SC 42.
- → **ISO/IEC 5059, Software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – Quality Model for AI-based systems**

Various Technical Reports give an overview of the current state of the art. These include:

- → **ISO/IEC TR 24027, Information technology – Artificial Intelligence (AI) – Bias in AI systems and AI aided decision making**
- → **ISO/IEC TR 24368, Information technology – Artificial intelligence – Overview of ethical and societal concerns.**

Projects on the following topics are currently being coordinated and are expected to start work in autumn 2020:

- → A certifiable management standard for AI that contains requirements and organizations for the responsible development and use of AI systems.
- → Various projects on the description of methods and processes for data quality in the context of machine learning. One of these projects is under the leadership of a German editor.

4.1.2 Requirements, challenges

The evaluation of AI applications with regard to their suitability can be based on ethical, legal and technical criteria. In view of the progressively growing AI market, an overview of application scenarios as well as embedded methods (4.1.2.1) and capabilities (4.1.2.1.2) of AI is indispensable. This helps to avoid shortcomings in development, deployment, conformity assessment and the determination of quality characteristics of AI. While 4.1.2.1.3 gives an overview of applications with embedded methods and capabilities of AI within software markets, a classification of AI applications based on different degrees of decision autonomy is presented in 4.1.2.1.4. Besides a description of AI through methods, capabilities and degree of autonomy, aspects such as “right to privacy, “basic right to life and physical integrity” can be reflected through criticality (4.1.2.1.5) (seeFigure 8).

Given the wide range of capabilities of AI applications, the potential for harm plays a decisive role in the social acceptance of AI. Using the example of AI-based recognition of traffic signs, the potential for harm can vary depending on the application: In road traffic, a substantial potential for harm can be assumed for self-propelled motor vehicles due to the high amount of concerns and obligations. In contrast, a conventional garbage truck with the same AI technology for traffic sign recognition does not represent a self-propelled vehicle, so that a minimal potential for harm can be assumed (see Figure 9).

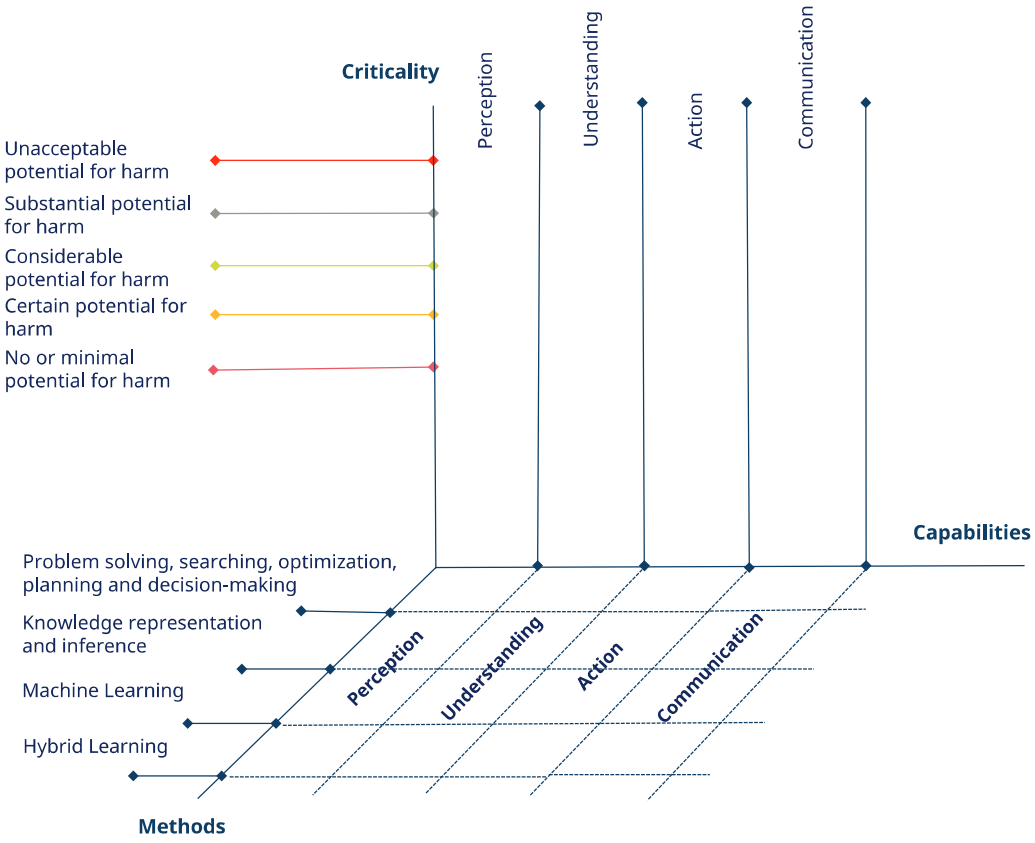


Figure 8: Three-dimensional classification scheme for evaluating an AI-based system

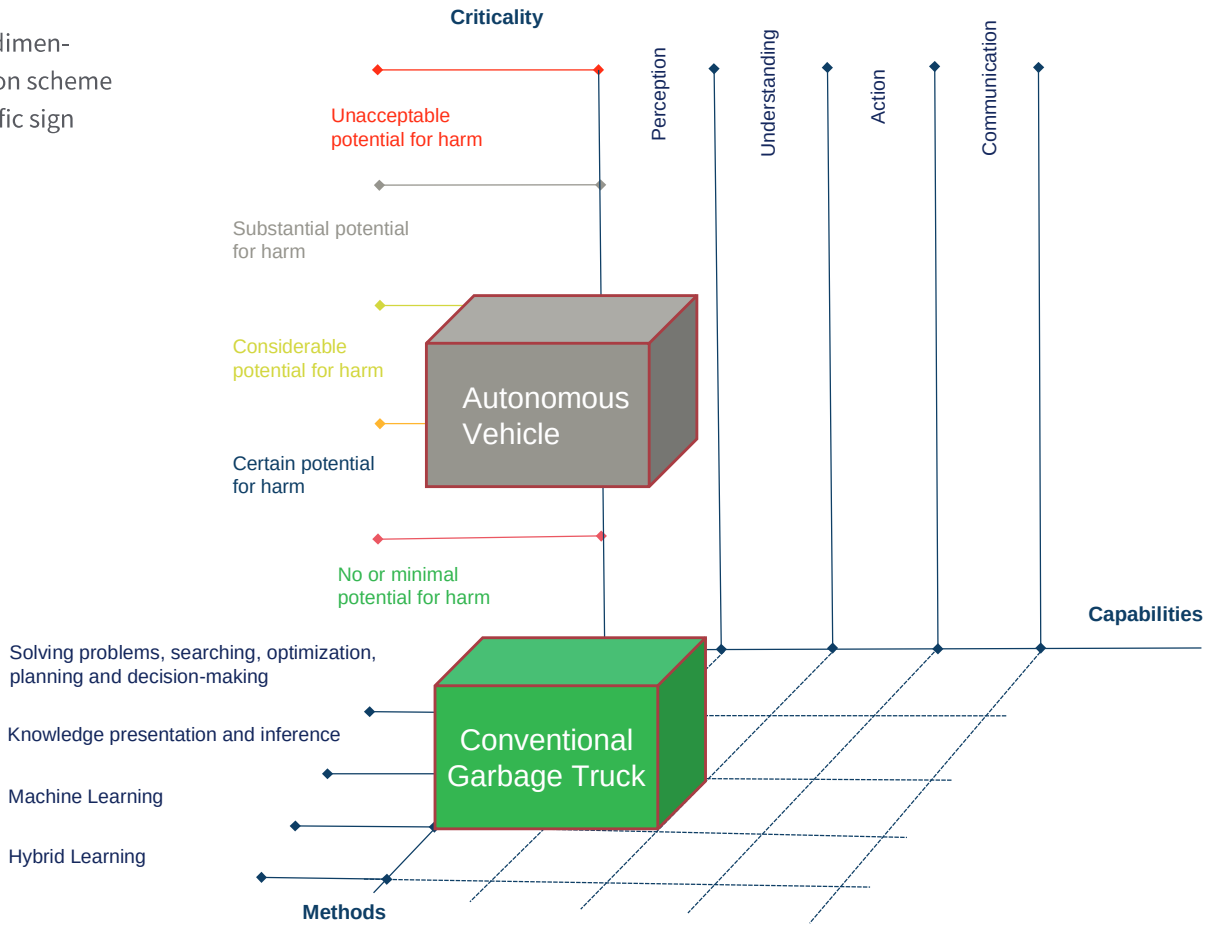


Figure 9: Three-dimensional classification scheme to evaluate AI traffic sign recognition

Figure 10: AI-based traffic sign recognition

A simple example of an AI application in a vehicle is a system for video-based traffic sign recognition. Here the detection of speed limits is already standard equipment in many cars. Since many traffic signs in connection with the permissible maximum speed have only a temporary validity (e.g. road works, gantries for dynamic traffic control), the necessary information cannot be taken from digital maps alone, but is recognized via pattern recognition using images from a camera, usually on the interior mirror. In this way, even recently erected signs, for example on construction sites, are registered. But that is not enough: Camera-based traffic sign recognition not only evaluates data based on signs. Instead, this recognition is compared with other assistance systems, such as the navigation system, the rain sensor and the time in order to correctly interpret restricted speed limits. However, driver assistance systems available on the market for traffic sign recognition do not work 100 % correctly, but a test showed a recognition rate between 32,5 % for the worst system and 95 % for the best system [35] on a course with 40 signs on speed limits which 12 cars passed. Temporarily invalidated speed signs using adhesive tape as well as speed displays in tunnels and illuminated signs on sign gantries proved to be a great challenge, as did the mix-up of speed limits for a turning lane (see Figure 10).

This simple example makes it clear that standards and test methods for this relatively simple subtask for autonomous driving according to Level 5 are necessary to ensure conformity of driving with road traffic regulations.

For this purpose, a standardized training data set for the traffic signs must be specified and benchmark tests must be provided for certification. In a risk-based approach, a detection rate of 99,9 % would have to be achieved for autonomous driving, while detection rates below 80 % also show considerable risks for product liability with pure assistance functions.

4.1.2.1 Classification

Following the position paper “A definition of AI: Main capabilities and scientific disciplines” of the AI HLEG [30], a distinction is made between methods and capabilities of AI. In both cases the following classifications are based on the standard work of Russell and Norvig [31] and integrate the current state of the art. The matrix in Table 2 shows which AI methods are used to realize certain AI capabilities. In order to also adequately reflect the actual state of the current industrial

AI markets, a classification of AI applications resulting from AI methods and AI capabilities is also carried out. Detailled information can be found in Table 2 to Table 5 and in the recently published Beuth Pocket [36].

4.1.2.1.1 Classification of AI methods

The methods of AI generally move within a kind of spectrum between symbolic and sub-symbolic – sometimes also called numerical – methods. In terms of symbolic methods, there are especially techniques of knowledge representation and logical reasoning, while sub-symbolic methods are primarily represented by techniques of machine learning. In between are methods of problem solving/optimizing/planning/decision-making as well as hybrid learning methods that use both symbolic and sub-symbolic techniques.

Symbolic AI is especially characterized by a deductive procedure, i.e. by the (algorithmic) application of logical rules or relations to individual cases. A distinction is made between methods for representing knowledge on the one hand and methods for applying this knowledge on the other. Knowledge can be represented either as certain or uncertain. In knowledge application the classical methods of logical reasoning are suitable for certain knowledge. For reasoning based on uncertain knowledge, probabilistic approaches are widely used, but there are also a number of nonprobabilistic approaches.

Subsymbolic AI is characterized in particular by an inductive procedure, i.e. by the (algorithmic) derivation of general rules or relationships from individual cases. In most cases a distinction is made between supervised learning to achieve a given goal and unsupervised learning without a comparable goal. When both approaches are combined, it is referred to as partially supervised learning. In addition, there is also known reinforcing learning without fixed target parameters, which does not require a fixed target value, but qualitative specifications (right/wrong).

The method complex of problem solving/optimizing/planning/decision-making comprises algorithms and procedures that focus on these sub-areas. Examples are intelligent agents, methods of game theory and evolutionary algorithms.

Hybrid procedures are often characterized by the fact that they combine sub-symbolic with other AI techniques, e.g. to be able to work both inductively and deductively. In contrast to classical sub-symbolic procedures, a form of knowledge representation is often used additionally. In contrast to classical symbolic methods, however, such knowledge representations are often algorithmically modified depending on input data.

Table 2: Classification of AI methods

CLASSIFICATION OF METHODS ACCORDING TO TOPICS			EXAMPLES
PROBLEM SOLVING, SEARCHING, OPTIMIZATION, PLANNING, DECISION-MAKING	Problem solving	Problem solving agents, problem solving through searching, search strategies	Uninformed and informed search strategies
			Adversarial searching (game theory)
			Searching with boundary and secondary condtions (constraint solving)
	Optimization	Statistical optimization methods	Local search for optimization
			Searching in continuous spaces
			Searching with partial observation
			Searching in unknown environments
			Dynamic programming
		Bio-inspired optimization methods	Evolutionary algorithms
			Genetic algorithms / genetic programming
			Swarm intelligence
	Planning and plan recognition	Autonomous and semi-automatic planning methods	State space search
			Planning graphs
			Hierarchical planning
			Planning in non-deterministic domains
			Time and resource planning methods
			Generation of plans
		Plan recognition methods	Plan recognition via abductive reasoning
			Deductive plan recognition
			Recognition via plan libraries
			Recognition via plan synthesis
	Decision-making	Approaches to decision-making	Models
			Use / value of information
			Decision networks
			Decision-theoretical expert systems
			Sequential decision problems
			Iteration models

	CLASSIFICATION OF METHODS ACCORDING TO TOPICS		EXAMPLES
KNOWLEDGE REPRESENTATION AND INFERENCE	Representation of knowledge	Knowledge representation languages and models	RDF
			RDFS
			OWL
			KIF
			Structure and formality
		Ontological engineering	Taxonomy
			Ontology
			Interpretation
			Calculus
			Deduction
			Abduction
			Ontology mapping
		Knowledge graphs and semantic networks	Knowledge networks / graphs
			Existence graph
			Graph traversing algorithms
			Mapping
			Semantic Web
		Modelling in formal logic	Propositional logic and predicate logic
			Higher-level logics, non-monotonic logics
			Temporal and modal logic
	Logical reasoning	Automatic proof methods	Resolution provers, connection provers
			SAT and SMT solvers
			Model checking
		Interactive proof methods	Tactical theorem proving
	Uncertain knowledge	Quantifying uncertainty	Bayes's rule
		Representation of uncertain knowledge	Bayesian network
	Probabilistic reasoning	Inference in Bayesian networks	Exact inference
			Approximate inference
			Markov chain simulation
		Relational probability models	Relational probability models in closed/open universes
		Time and uncertainty in probabilistic reasoning	Hidden Markov model
			Kalman filter
			Dynamic Bayesian networks
	Non-probabilistic approaches	Qualitative approaches	Reasoning with default information
			Truth Maintenance Systems (TMS)
		Rule-based approaches	Rule-based reasoning with "certainty factor"
		Vagueness reasoning	Fuzzy quantities and fuzzy logic
		Reasoning with belief function	Dempster-Shafer theory
	Further approaches to uncertain reasoning		Spatial reasoning
			Case-based reasoning
			Qualitative physics
			Psychological reasoning

	CLASSIFICATION OF METHODS ACCORDING TO TOPICS		EXAMPLES
MACHINE LEARNING	Supervised learning	Neural networks	Multi-layer perceptron
			Learning Vector Quantization (LVQ)
			Radial basis function networks (RBF)
			Adaptive Resonance Theory (ART)
			Convolutional Neuronal Networks (CNN)
			Recurrent Neural Networks (RNN)
			Time Delay Networks (TDNN)
			Long-Short Term Memory (LSTM)
			Hopfield networks
			Boltzmann machines
		Statistical learning	Decision trees
			Random Forest
			Support Vector Machine (SVM)
		Probabilistic methods	Naive-Bayes
			Fuzzy Classifier
	Unsupervised learning	Clustering	k-means
			Hierarchical clustering
			DBSCAN
			Fuzzy clustering
			Self-organizing map
		Dimension reduction	Autoencoder
			Principal component analysis
		Probabilistic methods	Fuzzy k-means
	Partially supervised learning	Statistical approaches	Expected Value Maximization (EM) with generative mix models
			Transductive Support Vector Machines
		Modified learning strategies	Self-training
			Co-training
		Graph-based approaches	Graph-based approaches
	Reinforcement learning	Temporal Difference Learning	Q-Learning
			SARSA
		Monte-Carlo methods	Markov Chain Monte Carlo
		Adaptive dynamic programming	Active and passive adaptive dynamic programming
HYBRID LEARNING METHODS	Hybrid neural systems	Unified Neural Architectures	Constructivist Machine Learning
		Transformation Architectures	Rule extraction for neural networks, neuro-fuzzy expert systems
		Hybrid Modular Architectures	
	Learning via knowledge structures	Logical learning	Current best learning
		Inductive logical programming	Sequential covering algorithm, constructive induction algorithms
		Explanation-based learning	
		Learning using relevant information	
	Conversational learning	Active, dialogue-based learning	Dialogue-based supervised learning
			Dialogue-based reinforcement learning

4.1.2.1.2 Classification of AI capabilities

AI as a scientific discipline is inspired by human cognitive capabilities [31]. Such capabilities have been classified within didactics and pedagogy since the middle of the last century on the basis of so-called learning goals. The most widespread classification system in use today distinguishes human capabilities both in terms of six cognitive levels and four basic cognitive domains [37], which can be used to distinguish up to 24 human cognitive capabilities.

Against this background, all currently existing AI-based systems represent only a part of the human cognitive capability spectrum. If one follows the assumption that AI capabilities imitate human capabilities, they can be roughly divided into the core areas of perception, understanding, action and communication. Most of these capabilities are realized by combining mechatronic and software components. The proposed classification helps to structure the discussion, but is not selective.

AI capabilities from the field of perception include information processing through the sensory abilities of image understanding, sound interpretation, haptics, smell and taste processing up to the complex field of recognition and interpretation of social signals.

The capability to understand is used to describe information processing in terms of evaluation, prediction and decision-making. The spectrum includes the sub-items fusion of perceptions, episodic memory, explanation and self-regulation.

The AI capability action describes in particular mechanically or physically executed activities such as robot perception, motion planning, sensor technology and manipulators, kinematics and dynamics, as well as the field of human-robot interaction, since this form of interaction focuses on physical human-machine interaction.

Communication refers to the transmission of information for processing natural language and during human-machine interaction. In computational linguistics, natural language processing corresponds to the skills of text generation, machine translation, text analysis, information and knowledge extraction, information retrieval, document analysis and speech dialogue systems. Human-machine interaction involves cognitive systems and interaction paradigms and modalities.

Table 3: Classification of AI capabilities

CAPABILITIES OF ARTIFICIAL INTELLIGENCE			EXAMPLES
PERCEPTION	Sensor data processing and interpretation	Image understanding	Image analysis, object recognition, video analysis, perceptual reasoning, scene analysis, photometry, physical attributes, 3D modelling, simulation, virtual reality
		Noise interpretation	Language recognition and synthesis, noise recognition and synthesis, anomaly recognition
		Haptics	Near-sensor technologies and methods of perception for tactile input and output (sensations like structure, tickling, touch, movement, vibration, temperature, pressure and tension)
		Social signals	Recognition and interpretation of gestures, facial expressions, body posture, affects and mood, emotions
		Smell and taste	Near-sensor technologies and methods of perception to recognize and synthesize smells, recognition of smell anomalies, and recognizing taste

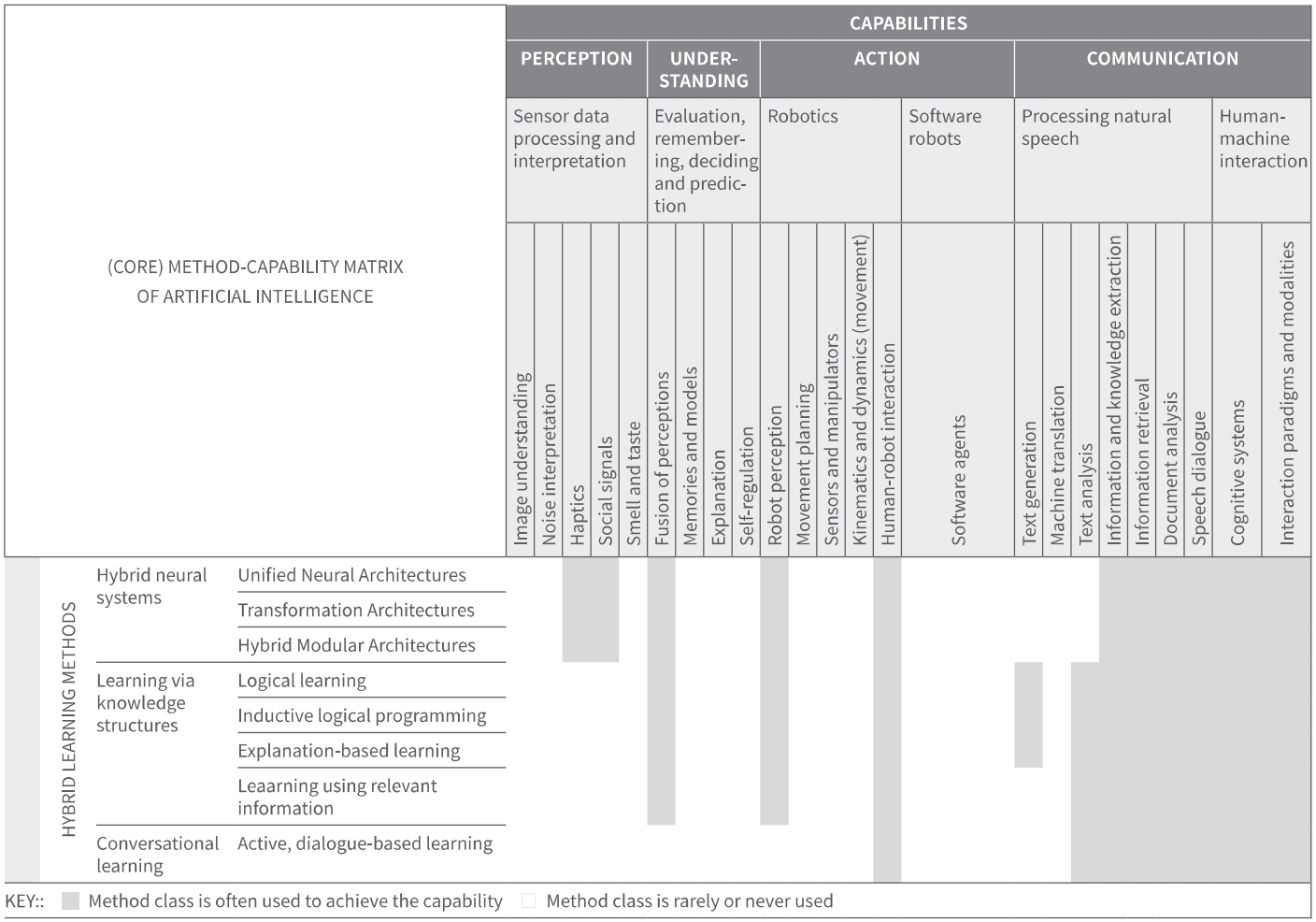
CAPABILITIES OF ARTIFICIAL INTELLIGENCE			EXAMPLES
UNDERSTANDING	remembering, deciding and prediction	Fusion of perceptions	Sensor data fusion and interpretation at the semantic level, data association, decision fusion, status assessment, ML-based/model-supported/factorgraph-based/probabilistic sensor data fusion methods
		Memories and models	Episodic and semantic memory, task and process modelling, environment modelling, process memory, discourse memory, plan library
		Explanation	Explanation derivation and generation, rationalization, hybrid models, integrated prediction and explanation models, explanation through architecture modification, model-diagnostic explanation
		Self-regulation	Modelling own performance limits, resource-adaptive action planning, methods of self-optimization, dynamic "world modelling"
ACTION	Robotics Software robots	Robot perception	Near-sensor technologies and methods of perception in robot systems, 2D and 3D perception methods, localization
		Movement planning	Methods of planning unsure movements, control methods
		Sensors and manipulators	Passive and active sensors, effectors, manipulators, cooperating manipulation
		Kinematics and dynamics (movement)	Kinematics systems, spatial kinematics, forwards kinematics, inverse kinematics, dynamic movement systems
		Human-robot interaction	Soft robotics, human-robot collaboration, multi-modal teleoperation
		Software agents	"Autonomous software systems, process automatization, (Chat-)Bots that carry out transactions, acting assistance systems"
COMMUNICATION	Processing natural speech	Text generation	Paraphrasing, Markov text generation, meaning-text model, generation of relationships, reports, artistic texts
		Machine translation	Transfer and interlingual methods, example-based, static, neural and semi-automatic approaches
		Text analysis	Parsing (syntactic analysis), shallow and deep analysis (semantic interpretation)
		Information and knowledge extraction	Text and web mining, entity extraction, disambiguation, relation extraction, event extraction
		Information retrieval	Vector space model, LSA, pLSA, semantic search, fact search, question-answer systems, autocomplete
		Document analysis	OCR, ICSR, document classification, segmentation, range recognition
		Speech dialogue systems	Speech act recognition, reference resolution, explanation dialogue, discourse modelling, dialogue management, language change strategies
	Human-machine interaction	Cognitive systems	Human factors, human processor models, user modelling, cognition theory (cognition, mental models, memory, learning type, cognitive load)
		Interaction paradigms and modalities	Interaction design, patterns, multimodal interaction, user experience, fusion and fission of modalities

Using the classification matrix for methods and capabilities, a labelling requirement for implemented methods and capabilities can be established for AI applications. Chapter 4.3 provides an overview of requirements and challenges regarding the conformity assessment and quality assessment of AI-based systems.

Table 4: Method-capability matrix

(CORE) METHOD-CAPABILITY MATRIX OF ARTIFICIAL INTELLIGENCE			CAPABILITIES																								
			PERCEPTION					UNDER-STANDING				ACTION					COMMUNICATION										
			Sensor data processing and interpretation					Evaluation, remembering, deciding and prediction				Robotics				Software robots	Processing natural speech							Human-machine interaction			
			Image understanding	Noise interpretation	Haptics	Social signals	Smell and taste	Fusion of perceptions	Memories and models	Explanation	Self-regulation	Robot perception	Movement planning	Sensors and manipulators	Kinematics and dynamics (movement)	Human-robot interaction	Software agents	Text generation	Machine translation	Text analysis	Information and knowledge extraction	Information retrieval	Document analysis	Speech dialogue	Cognitive systems	Interaction paradigms and modalities
METHODS	PROBLEM SOLVING, SEARCHING, OPTIMIZATION, PLANNING, DECISION-MAKING	Problem solving	Problem-solving agents, problem solving through searching, search strategies														To be taken from the previous columns according to the application										
		Optimization	Statistical optimization methods																								
			Bio-inspired optimization methods																								
		Planning and plan recognition	Autonomous and semi-automatic planning methods																								
			Plan Recognition Methods																								
		Decision-making	Approaches for Decision Making																								

(CORE) METHOD-CAPABILITY MATRIX OF ARTIFICIAL INTELLIGENCE			CAPABILITIES																						
			PERCEPTION					UNDER-STANDING				ACTION						COMMUNICATION							
			Sensor data processing and interpretation					Evaluation, remembering, deciding and prediction				Robotics					Software robots	Processing natural speech						Human-machine interaction	
			Image understanding	Noise interpretation	Haptics	Social signals	Smell and taste	Fusion of perceptions	Memories and models	Explanation	Self-regulation	Robot perception	Movement planning	Sensors and manipulators	Kinematics and dynamics (movement)	Human-robot interaction	Software agents	Text generation	Machine translation	Text analysis	Information and knowledge extraction	Information retrieval	Document analysis	Speech dialogue	Cognitive systems	Interaction paradigms and modalities
KNOWLEDGE REPRESENTATION AND INFERENCE	Representation of knowledge	Knowledge representation languages and models																							
		Ontological engineering																							
		Knowledge graphs and semantic networks																							
		Modelling in formal logic																							
	Logical reasoning	Automatic proof methods																							
		Interactive proof methods																							
	Uncertain knowledge	Quantifying uncertainty																							
		Representation of uncertain knowledge																							
	Probabilistic reasoning	Inference in Bayesian networks																							
		Relational probability models																							
		Time and uncertainty in probabilistic reasoning																							
	Non-probabilistic approaches	Qualitative approaches																							
		Rule-based approaches																							
		Reasoning with vagueness																							
		Reasoning with belief function																							
	Further approaches to uncertain reasoning																								
MACHINE LEARNING	Supervised learning	Neural networks																							
		Statistical learning																							
		Probabilistic methods																							
	Unsupervised learning	Clustering																							
		Dimension reduction																							
		Probabilistic methods																							
	Partially supervised learning	Statistical approaches																							
		Modified learning strategies																							
		Graph-based approaches																							
	Reinforcement learning	Temporal Difference Learning																							
		Monte Carlo methods																							
		Adaptive dynamic programming																							



4.1.2.1.3 Classification of AI applications

The classification of AI applications is often based on the AI methods and AI capabilities described above. The aim of the AI application is to concretely implement mathematical methods and abstract capabilities using software. In this way, specialized software markets have emerged to market these typical AI products. These can be purchased or rented by companies and users to increase the productivity of business processes or to enable innovations in business models. In addition, the typical software markets (see Table 5) are uniformly designated worldwide and are regularly monitored by independent market analysts (e.g. IDC, Gartner, Forrester, etc.), so that potential users, projects and investors are well informed about the status of capabilities.

The software markets can be roughly divided into business intelligence & decision support, AI-based customer interaction, AI- based services and AI development environment & tools.

Business intelligence & decision support focuses on the timely and topic-oriented creation of reports. These are designed to provide a quantitative and qualitative overview of the business and have been commercially available for many years in all areas – e.g. finance, human resources (HR), development, marketing and sales. This supports decisions and enables complete planning processes in complex environments. These capabilities include analytics, as they typically require the analysis of multidimensional data spaces. Key products in this area are software environments for mathematical and AI-based optimization and the calculation of forecasts. Another area is the processing of speech typically used for searching, navigation and exploration in large text bodies. When several of these functions are combined, entire business processes can be automated, often referred to as Robotic Process Automation (RPA).

Since 2012 the AI trend has accelerated considerably due to the fact that the available CPUs and GPUs (central and graphics processing units) are becoming more and more powerful

and AI methods based on artificial neural networks can be realized faster and cheaper. This allows new possibilities for the human-machine interface based on AI applications that simulate SMS, chats, speech and physical movements and automate corresponding processes, for example simple dialogues in call centres and service centres.

To simplify the use of AI applications, typical AI applications are offered in public or private cloud environments. This allows the user to start immediately with the adaptation of the application to their own needs without having to spend a lot of time and effort on building hardware and software. Typical AI services that are offered out-of-the-box are: image recognition, video analysis, speech-to-text conversion, text-to-speech conversion, translation, text analysis, intelligent search and machine learning. In all of them the actual use of the artificial neural network is encapsulated and facilitated by a simple graphical user interface or by simple function calls from standard languages (e.g. Java, C, Python, etc.).

Appropriate AI development environments and tools are needed for the development of AI applications. These take into account the typical phases of an AI project: Build, Train und Run. In all phases, open source technologies and software libraries are frequently used, which on the one hand offer AI methods and on the other hand professional software development, e.g. method-based and in distributed teams.

By regulating systems based on AI, possible inadequacies of AI applications and competition-distorting constellations can be avoided. In line with the European Commission's White Paper "On Artificial Intelligence – A European Approach to Excellence and Trust", the following aspects are important with regard to regulation: liability, transparency and accountability, as well as training data, retention of data and records, information to be provided, robustness, accuracy, human oversight and specific requirements for certain AI applications, e.g. remote biometric identification applications.

The ethical aspects of the development, benefits and standardization of AI are currently under special discussion. Here, an important role is played by the following characteristics, which should be methodically and technically thought through and ensured for each AI application: autonomy & control, transparency, stability against disturbances, security and all aspects of data protection.

Table 5: Overview of software markets and typical AI applications

Software markets & typical AI applications		
Software market	**Typical software products**	**Principles**
Business Intelligence & Decision Support Systems	Business Intelligence	Autonomy & Control
	Decision Support	
	Workflow systems	
	Planning Analytics	
	Constraint Based Optimization	
	Prediction Capability	Fairness
	Text Processing Platforms & Search Engines	
	Robotic Process Automation (Rule-Based)	
	Cognitive Automation (Training-Based)	
	Real-Time Processing	

Software markets & typical AI applications		
Software market	**Typical software products**	**Principles**
AI based Customer Interaction	Chatbots	Transparency
	Voicebots	
	Avatars	
	Virtual & Augmented Reality	
AI based Services consumed from Public or Private Cloud	Image Recognition	
	Video Analytics	Robustness
	Speech To Text	
	Text To Speech	
	Translation	
	Deep Learning as a Service	
	Knowledge Navigation	Security
	Knowledge Exploration	
	Intelligent Search	
	Natural Language Processing	
	Automatical Annotation	
AI Development Environment & Tools	Build & Develop AI	Data Governance
	Train & Optimize AI	
	Run & Manage AI	
	Ethic Support Tools	

4.1.2.1.4 Classification of AI autonomy

AI applications and the computer systems that implement them can have different degrees of decision autonomy [33]. For example, the Data Ethics Commission of the German Federal Government [10] distinguishes three classes of autonomy: algorithmically-based, algorithmically-driven and algorithmically-determined systems.

Algorithmically based AI applications work as pure assistance systems without autonomous decision-making authority. However, the (partial) results and (partial) information calculated by them are the basis of human decisions.

Algorithm-driven AI applications take partial decisions from humans or shape human decisions through the results they calculate. As a result, the actual decision-making scope of humans and consequently their possibilities for self-determination shrink.

Algorithmically-determined AI applications make decisions independently and thus exhibit a high degree of autonomy. Due to the high degree of automation, there is no longer a human decision in individual cases, especially no human review of automated decisions.

4.1.2.1.5 Risk-based assessment of applications

In view of the diversity, complexity and dynamics of applications, the Data Ethics Commission sees a need for risk-based assessment. The aim is to contribute to a human-centred and value-oriented design and use of systems. Against this background, on the basis of an ethical-legal regulatory framework, specifications for transparency, explainability and traceability are planned. Special emphasis will be placed on the aspects of the scope of information rights and obligations, as well as liability by human decision-makers.

The assessment is intended to be based on a criticality pyramid. According to the pyramid, a possible occurrence of damage (e.g. human-induced and/or algorithmically determined) is to be assessed with its extent (e.g. “right to privacy, fundamental right to life and physical integrity” and “prohibition of discrimination”) for a socio-technical system. For the assessment, the involvement of all technical components (including hardware, software and training data), human actors (including developers, manufacturers, testers and users) and life cycle phases (including development, implementation, conformity assessment and application) is sought. In addition to the legislator, developers, testers and users should also be able to assess the criticality of a system using the pyramid.

The criticality pyramid (see Figure 11) has five levels (or degrees) of criticality. As the level of criticality increases, the demands on a socio-technical system to be evaluated grow. Level 1 systems: “Applications without or with only minimal potential for harm” are checked for quality requirements and are not subject to a risk-based assessment (application example: automatic purchase recommendation; anomaly detection in industrial production). A risk impact assessment should be carried out for Level 2 to Level 5 systems. Level 2 systems, “applications with a certain potential for harm”, should have disclosure requirements on transparency. In addition, investigations into misconduct are necessary, for example by analysing the input and output behaviour (application example: non-personalized, dynamic pricing; automatic processing of claims settlement). For Level 3 systems “applications with regular or considerable potential for harm”, approval procedures should be used in addition to the measures at Level 2 (application example: automatic credit allocation; fully automated logistics). Level 4 systems “applications with substantial potential for harm” should, in addition to the measures of Levels 2 and 3, fulfil further obligations for control and transparency, such as publication of algorithms and calculation parameters, as well as the creation of an interface to directly influence the system (application example: AI-based diagnostics in medicine; automated driving). Systems at Level 5, “applications with unacceptable potential for harm” shall be subject to a proportionate or complete ban on use (application example: systems that override the presumption of innocence, or systems that have an approvingly lethal effect without human influence).

With regard to AI, the application of the criticality pyramid has revealed a further, more profound need for discussion. In the course of this, a procedure for the legal assessment and the ethical evaluation of AI applications should crystallize. This would make it possible, for example, to define the scope of basic and liability rights for an AI application. Furthermore, the significance of the criticality pyramid could be increased by including several additional dimensions, so that a possible extent of harm can be described more concretely. In addition, certification in the course of a conformity assessment should be able to demonstrate the fulfilment of requirements with regard to the potential for harm of AI applications within Levels 1 to 4. For Level 5, the demonstration of conformity is to be prohibited, since, for example, the prevention of a high level of harm cannot be ensured through certification. In conclusion, there is the greatest variety of obligations, requirements, reservations, concerns, ethical and legal implications with regard to regulation and conformity assessment certification for systems at Levels 2 to 4.

For the assessment of AI-relevant criteria, standardized conformity assessment procedures of accredited testing laboratories can be used, for example based on the ISO/IEC 17000

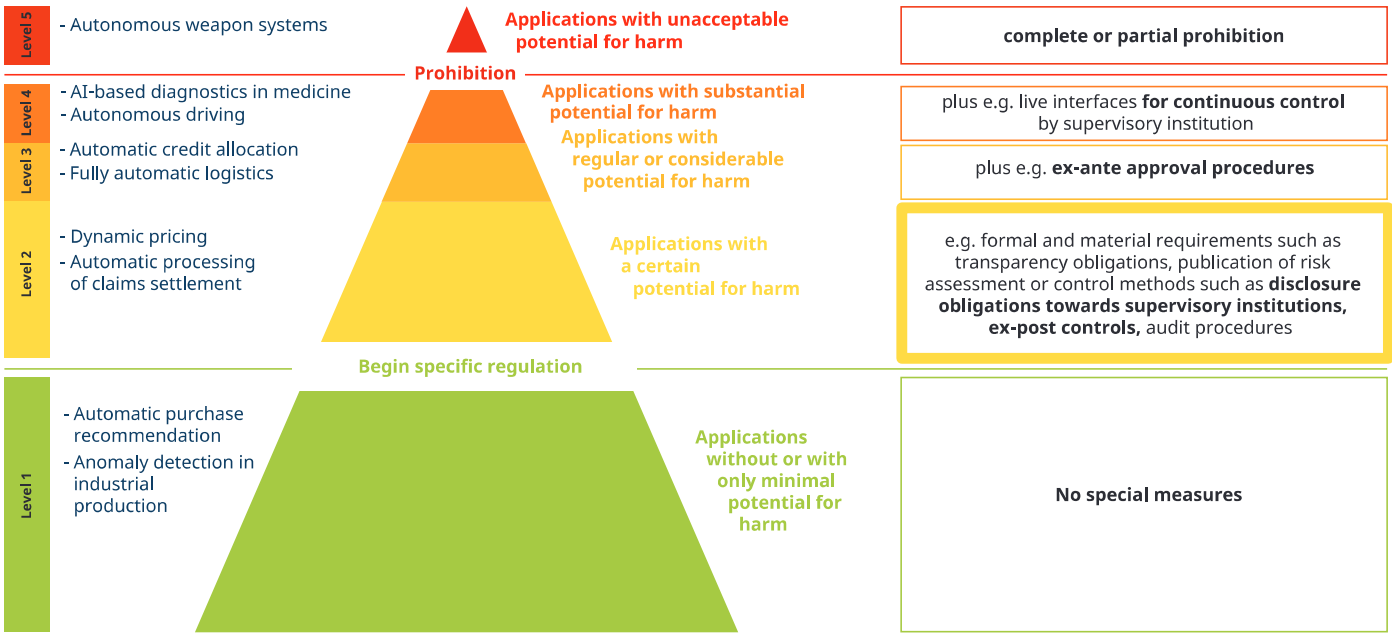


Figure 11: The criticality pyramid [10] and a risk-adapted regulatory system for the use of algorithmic systems

series of standards [38]–[44]. In the course of conformity assessment, products, systems and processes may be subject to testing, calibration, inspection or certification, and persons to certification. To this end, the expertise of already established, accredited certification bodies should be expanded with regard to the methods and capabilities of AI. An insight into relevant aspects of conformity assessment with a focus on AI is provided in Chapter 4.3.

4.1.2.2 Trustworthiness

The term "trustworthiness" can basically refer to both organizations and technical systems. A technical system (i.e. a product or an electronically provided service) can be trusted with regard to certain properties such as security or reliability if there is evidence (e.g. in the form of a test report or a certificate) that the system meets such properties.[17] The trustworthiness of an organization is broader: It refers to an organization being trusted to implement appropriate measures and maintain management structures – a management system – to meet the expectations of its shareholders and other interested parties. In addition to a corresponding test report, the reputation of an organization or its acceptance in the market can also contribute to its trustworthiness[17].

In the context of this paper, technical systems that implement AI functions (called AI systems), or organizations that implement, offer or operate such systems will be considered.

4.1.2.2.1 Requirements for trustworthiness

In its ethical guidelines [5] the High-Level Expert Group on Artificial Intelligence (AI HLEG) has described a number of requirements for AI systems with regard to their trustworthiness. In most cases, these are hybrid applications, i.e. they consist of AI components and not AI-based software and hardware, and are basically understood as special IT. In this chapter, these requirements will be used as representatives for a number of similar approaches to derive standardization needs. Figure 12 gives an overview of the requirements mentioned in the guidelines, which are further discussed below:

1. **Priority is given to human agency and oversight**, and the observance and safeguarding of fundamental rights are also mentioned. It is required that information, supervision and control mechanisms are available in connection with AI systems in order to avoid negative effects, e.g. on basic rights, but also the misuse of AI systems. On the one hand, these questions have technical impli-

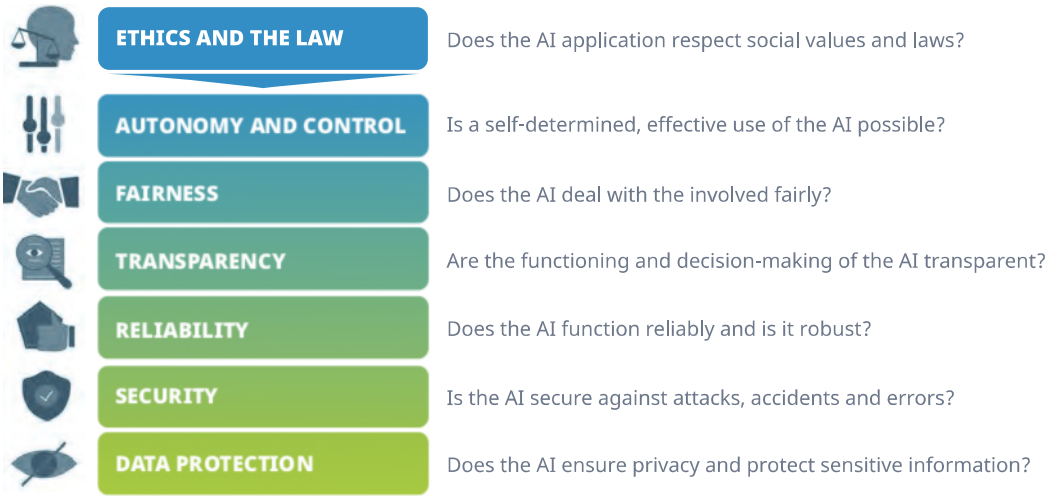


Figure 12: Requirements for a trustworthy AI [along the lines of [45]]

17 Ultimately, here the trustworthiness of a technical system is attributed to the trustworthiness of an organization, namely the testing body. However, since the audit refers to the system and not to its manufacturer or provider, this distinction between system trustworthiness and organizational trustworthiness should be maintained to better structure the discussion.

cations that relate to the development of AI systems, namely the implementation of effective monitoring and control functions. However, the use of such functions must be embedded in the management processes of the operating organization in order to be effective. After all, the question of the process of human action and control of technical systems by humans refers to the objectives, the mission and the willingness to take risks of an organization operating AI systems (governance). In the context of public security, for example, different considerations will play a role for the use of AI than for use by a commercial enterprise. The AI HLEG demands that an impact assessment be carried out in areas where the use of AI may affect fundamental rights.

2. **Technical robustness and safety**, e.g. resiliance to attacks and security breaches, fall back plan and general safety, accuracy, reliability and reproducibility. From the perspective of standardization, an entire range of relevant questions arise:

→ Are common approaches to management IT or cybersecurity sufficient for the use of AI? What are the specific vulnerabilities of AI systems? Are new controls or management processes necessary?

→ What restrictions must an AI system be subject to? When does the AI have to be restricted or overruled by classical systems or by humans in order to avoid damage to persons or objects?

→ How can the precision of AI systems and their reliability be measured or ensured? Are there generally accepted metrics and units of measurement? What role do development and quality assurance processes play?

1. **Privacy and data governance**, such as respect for privacy, data quality and integrity, and data access. Questions concerning standardization activities are data protection management in connection with AI, but also how data quality can be ensured overall. This applies in particular to the case where data for machine learning is provided by external providers.
2. **Transparency**, e.g. traceability, explainability and communication. On the one hand, the AI HLEG requires that data records and processes that led to the decision of the AI system be documented. On the other hand, the term "explainability" refers to the traceability of the internal function of AI systems (e.g., the question with which criteria an automatic decision was made by an AI system).
3. **Diversity, non-discrimination and fairness**, e.g. avoiding unfair bias, accessibility and universal design and stakeholder participation.
4. **Societal and environmental well-being**, e.g. sustainability and environmental friendliness, social impact, society and democracy.
5. **Accountability**, e.g. auditability, minimizing and reporting negative impacts, trade-offs and redress.

In summary, let it be said that the AI HLEG recommendations address a number of important issues. However, the publication cannot be used directly to derive mandates to the standardization bodies:

1. Standards are basically of a technical nature, i.e. they refer to requirements and recommendations of a technical-organizational nature and how such can be applied within an organization. Social, legal and political requirements cannot be codified in standards, only technical-organizational implications resulting from such requirements can become the subject of a standard Thus, not all topics mentioned by the AI HLEG are already suitable for standardization.
2. The AI HLEG does not distinguish between trust in the AI product or service (in the sense of a product or service that uses AI functions), and trust in the organization that provides such a service or uses, manufactures or distributes such a product.
3. If standardization is seen as an objective at international level, i.e. within ISO, IEC or the ITU, an ethical basis for such work must be dispensed with unless it is generally accepted in the international community. For example, the project to propagate a framework of values that is not internationally recognized with the help of an international standard is excluded by the principles of the World Trade Organization that are binding for these three organizations [46].

4.1.2.2.2 Trust in products and services

Common Criteria (CC)

The Common Criteria (CC) [47] describe a methodology for testing products and services with a focus on their security, which can be used as a conceptual framework for corresponding tests of AI systems. The CC are also available as an International Standard ISO/IEC 15408 [48]–[50]. A coordinated methodology for evaluation based on the CC is described in the International Standard ISO/IEC 18045 [51]. These documents form the technical basis of the Common Criteria Recognition Arrangement (CCRA) [52], which has been signed by a large number of countries, including Germany. Further information on the CC can be found on the website of the Federal Office for Information Security (BSI) [53], for example.

Requirements for testing according to the CC are summarized in the Evaluation Assurance Levels (EALs):

EAL1 functionally tested
EAL2 structurally tested
EAL3 methodically tested and checked
EAL4 methodically designed, tested and reviewed
EAL5 semi-formally designed and tested
EAL6 semi-formally verified design and tested
EAL7 formally verified design and tested

Certification up to EAL4 is internationally recognized.

4.1.2.2.3 Trust in organizations

The relationship between governance, management and technical-organizational measures – management systems

For further investigation of the AI HLEG requirements on the trustworthiness of AIs, a conceptual digression will be undertaken to distinguish between the terms "governance" and "management", as is currently done in ISO/IEC 38500 [54] (see Figure 13). It should be noted that the term "management system" refers to all three levels discussed in the following, namely the governing body, the management, and concrete technical and organizational measures.

Governance

Governance refers to the general tasks and the objective of an organization, its self-image and the resulting values, and the culture of the organization that determines its actions. A central concept is that of a willingness to take risks. According to ISO/IEC 38500 [54] the governing body of an organization is responsible for the implementation of its accountability and due diligence obligations. Questions of liability are of particular relevance in connection with AI, since the possible degree of autonomy of AI raises the question of who is liable for errors and damages. Governance should take this into account, since the legal framework in this field is developing dynamically. The governing body sets requirements and establishes guidelines that must be implemented within the organization. The governing body is also responsible for establishing management structures (processes, roles, responsibilities) and providing adequate resources.

Management

The management of an organization translates the requirements and guidelines of the governing body into concrete processes, roles and responsibilities. Examples of management tasks include:

→ The identification and analysis of potential risks and the establishment of options for action based on the willingness of the organization to take risks.
→ The establishment of a data protection management system and processes to ensure sufficient data quality.
→ The introduction of security management for AI-based IT systems.
→ Effective management of the development and operation of AI systems.

Technical-organizational measures

This term covers all technical and organizational tools available to management to fulfil their tasks effectively and verifiably. Technical-organizational measures range from the availability of encryption functions to increase data security

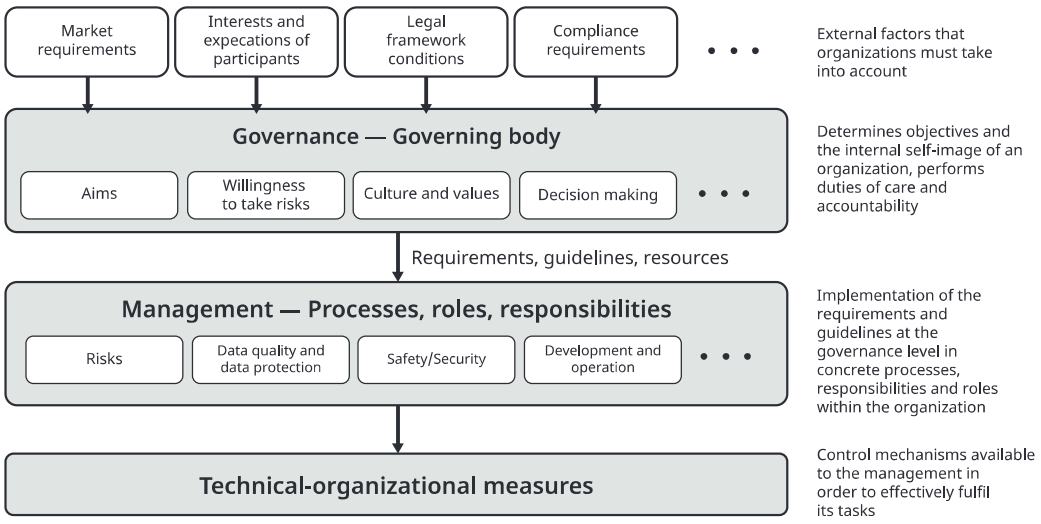


Figure 13: Management system: Governance, management and technical-organizational measures

to the application of statistical methods to identify unfair distortions or contamination in data sets and the availability of test and validation tools.

Requirements on the management system

The term management system standard (MSS) plays a central role in the context of international standardization. An MSS defines requirements for organizations for implementing effective and responsible management. In some cases, requirements are also placed on the governing body of an organization, and many MSS still contain specific controls in the sense of technical and organizational measures. The term "management system" thus refers to the overall picture presented in Figure 13. Minimum requirements for the management system are described in the Guidelines for International Standardization of ISO/IEC, in the so-called: High Level Structure (HLS) [55]:

1. **Context of the organization**; this includes, among other things, the legal framework, social expectations, needs and expectations of interested parties, goals and values of the organization, and the actual scope of the management system.
2. **Leadership**; the governing body must define binding readiness of the organization and lay it down in the form of guidelines. It must also define processes, roles and responsibilities for effective management.
3. **Planning**; this must describe activities to deal with risks and opportunities.
4. **Support**; this includes the provision of resources, the determination of necessary competencies, ensuring necessary mindfulness, communication and documentation.
5. **Operation**; this is the operational implementation of management requirements.
6. **Performance evaluation**; this comprises monitoring, analysis and evaluation, internal auditing and management review.
7. **Improvement**; this deals with the identification of non-conformity with regard to MSS requirements, corrective measures and the continuous improvement of the management system.

Organizations can demonstrate compliance with MSS (e.g. through self-assessment or certification by an independent third party), thereby increasing the organization's trustworthiness as regards the specific aspects of the MSS. When considering the use of a class of technologies such as AI, an organization's management system must therefore refer to the specific characteristics and range of impact of AI. This can be done by adding AI-specific requirements to existing MSS.

However, since the different MSS are published and maintained by different bodies in ISO and IEC, which have neither a common conceptual framework nor a synchronized way of working, and since it is not clear whether existing MSS are even sufficient to cover all aspects of AI, it is more promising to design a new MSS that focuses on AI-specific requirements.

Supporting specifications

MSS only include requirements for a management system, but do not describe its implementation. This allows organizations to define their own management structures in the way that suits them, as long as evidence can be provided that the MSS requirements are met. Such structures, but also underlying technical and organizational measures, are usually described in supplementary specifications, which now contain no requirements but only guidelines.

4.1.2.3 Development of AI systems

Software gives machines an ever-increasing range of functions. Hardware and software form a symbiosis and there are methods, such as V-Model® XT [56], [57] – with and without agile methods (e.g. Scrum) – which help ensure the quality of the overall result during development. For software with a predetermined functional sequence, there are generally accepted development and quality assurance procedures, such as code reading, module and application tests at various integration levels, verification and validation. These methods and procedures also work for software with rule-based AI systems. In addition to the quality of the software code and the compilers used, the software architecture, the quality of the data used and the learning phase are of particular importance when developing AI systems.

Learning AI systems receive essential functionalities through the learning phase. This learning phase can be static or dynamic, supervised or unsupervised. As with humans, the testing of what has been learned is a great and new challenge for software development. This is especially critical because AI systems show their strength especially where decisions or decision recommendations based on a large amount of data have to be made very promptly.

If AI systems are used for automated or autonomous decision-making in safety-critical areas, related procedures for verification and conformity assessment by third parties are also required. This applies in particular to evidence when proving functional safety in product liability.

An appropriate approach to the development of AI systems is a risk-based approach[18] considering the entire life cycle of an AI system in its application environment, as well as ensuring data quality in the learning and application phase.

Further consideration must be given to AI systems whose source code and/or learning content was generated by themselves or by other AI systems. Thus, an existing AI system develops a new one or changes its learning content, so that a kind of evolution of the machines takes place.

4.1.2.3.1 The life cycle of an AI system

Similar to traditional software development, the life cycle phases of an AI system consist of: **Concept, Development, Deployment, Operations and Retirement**, whereby especially for systems based on machine learning, which can be applied in different phases of the life cycle from development to operation, there is a much closer interlocking of the phases than is the case with classical software systems.

During the concept phase it has to be defined whether the application to be created is created as a rule-based, static or dynamic AI module and which requirements result from the context of the application area, as well as the necessary data quality. For rule-based AI systems, the established software life cycle according to ISO/IEC/IEEE 12207 [58], or for safety-critical systems also according to ISO 26262 [59]–[70], ISO/IEC 27034 [71]–[78] or IEC 61508 [79]–[86], can be applied. A risk-based approach is necessary for static and dynamic AI systems.

18 "Risk-based" in English.

Based on this, a risk analysis must be carried out, e.g. based on an FMECA (Failure Mode and Effects and Criticality Analysis), which must consider the entire life cycle of the AI system. As part of the risk assessment, a first simple classification, as presented in DIN SPEC 92001-1 [87], can be made. (see Figure 14). The separation into low risk and high risk can be sufficient, but a more fine-grained phase model seems to be more appropriate, especially since aspects of dynamic models can be dealt with in more detail.

As an alternative to the previously described DIN SPEC 92001-1 [87], VDE/DKE presented a "Reference Model AI" [88] (see Figure 15) which describes a development process for AI systems based on the V-Model® XT. A consensus model for the AI life cycle is to be developed within the framework of the Standardization Roadmap AI.

4.1.2.3.2 Data quality principles for AI modules

The quality of the data for learning, testing and subsequent application is an essential factor for the successful development and, in the application phase, for the use of AI systems. A general definition of data quality in software development is described in ISO/IEC 25012:2008 [89] and consists of inherent and system-dependent characteristics. To what extent this standard is also suitable for the development of AI applications, or if other or further quality features are important, has to be checked and, where necessary, standardized specifically for AI applications. It will make sense to tailor and/or prioritize the dimensions for the respective use case. If, for example, simulation data ("synthetic data") are used for learning and/or testing, their usability/exemplary nature must be ensured. If incorrect data is deliberately provided for learning, testing and inspection purposes, it must be marked

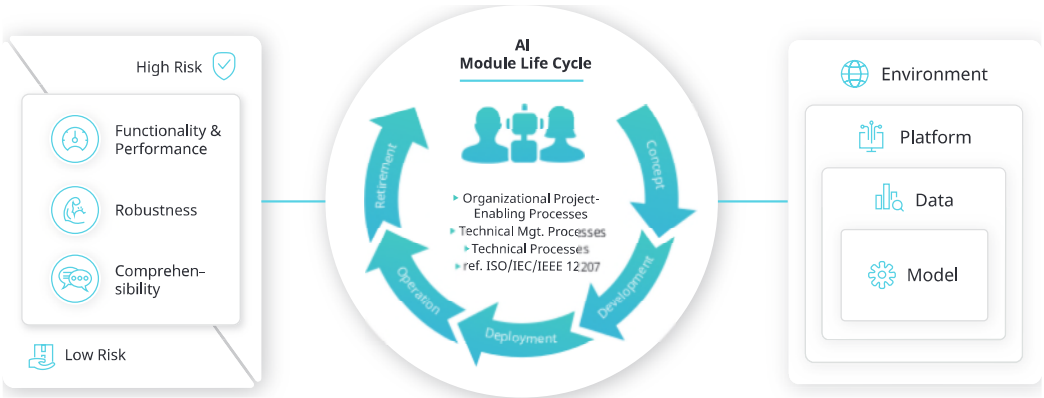


Figure 14: AI quality meta-model of DIN SPEC 92001-1 [87]

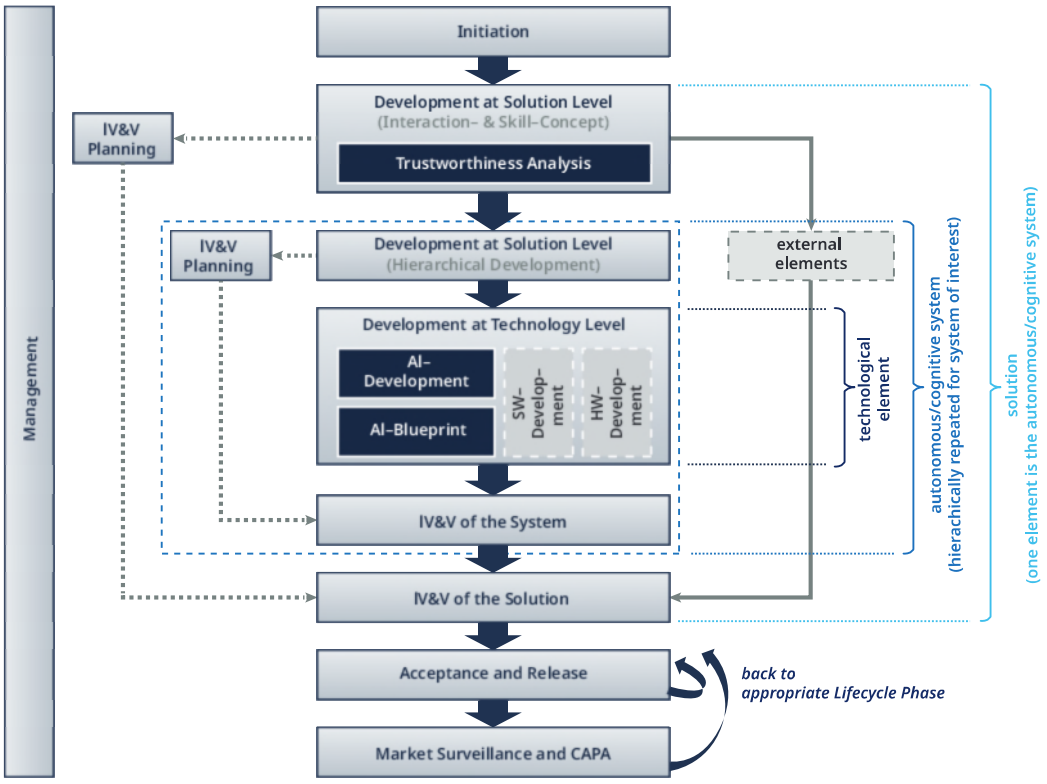


Figure 15: VDE/DKE Reference model AI [88]

accordingly and separated from the non-erroneous data in a suitable manner so that no unintentional mixing occurs.

The Fraunhofer Guidelines for High Quality Data and Metadata (NQDM) of 2019 [90] lists the following dimensions of data quality:

1. **Currency**
   Data describe the current reality. Therefore, it is recommended to pay attention to a time stamp and, if necessary, a version number when recording and naming the data. Data should be checked at appropriate intervals to ensure that they are representative.
2. **Accuracy**
   The data should contain correct values and be as error-free as possible. Here a datum is faulty if it does not correspond to its classification. Thus an incorrect datum, which has been communicated to the AI system for training as incorrect, is not incorrect in this sense. For the training of AI systems, incorrect data is deliberately used, but it is also classified as faulty.
3. **Precision**
   Depending on the application, the precision of the data is of high relevance, so that, for example, rounding of values should be avoided. The content descriptions of the data should also be as precise as possible in order to quickly assess the relevance of data.
4. **Conformity**
   When providing data, attention must be paid to the expectation conformity of the contained information in a certain usage context and format, for example when naming attributes and vocabulary. For a universal use of the data, appropriate standards should be used where possible, e.g. ISO 8601 [91] for dates.
5. **Consistency**
   Data should be free of contradictions, both in itself and across data sets. This dimension may already be covered by accuracy.
6. **Transparency and trustworthiness**
   The origin, originality and changes to the data should be made traceable, so that the transparency and credibility of the data can be strengthened, thereby gaining the trust of the users and also meeting ethical requirements.
7. **Reliability**
   In order to assess the reliability, or the degree of maturity, of a piece of information, it can be assigned a status (see also DCAT-AP.de).
8. **Understandability**
   The data structure, the naming of the data, as well as data interfaces should be easy to understand.
9. **Completeness**
   A data set should be complete: Attributes, which are mandatory for the further use of the data set, must therefore contain a value.