Lernende Systeme
GERMANY'S PLATFORM FOR ARTIFICIAL INTELLIGENCE

Hybrid AI

Combined use of knowledge and data

AT A GLANCE

Hybrid AI combines approaches based on knowledge representation and data-driven strategies (Machine Learning).

- **Credo:** What humans already know does not have to be learned from data.
- **Promise:** (Energy) efficient, robust, explainable and trustworthy AI systems that are subject to less bias, require less data for the learning process and are tolerant of erroneous or inaccurate data points.
- **Potential:** Applications that require reliable and accurate results (e.g. medicine, quality control).
- **Challenge:** Complexity and lack of guidelines for implementation as well as a lack of benchmarks.

Context

In the field of Artificial Intelligence (AI), Machine Learning (ML) with neural networks (Deep Learning) is currently attracting a great deal of attention. The reason for this are the impressive results that have recently been achieved in areas such as automatic text or image generation. These successes are essentially based on large amounts of data, the scalability of the models and high computing capacities. It is easy to get the impression that any goal can be achieved with Machine Learning if these three factors are strengthened. However, fundamental problems of large AI models cannot be solved by further scaling, including their high resource consumption.

In addition, it is often overlooked that some of the much-discussed successful applications are so-called hybrid AI systems. These combine the advantages of Machine Learning with human knowledge that is formally represented in various ways and processed by computers (table 3). If one of the approaches reaches its limits or achieves undesirable results, a combination can offer solutions.

Starting point for AI systems: Human knowledge

In order to create systems that automatically process information or even draw intelligent conclusions, existing knowledge can be used as a starting point – be it in the form of experience, expert knowledge or knowledge gained from research.

Knowledge and rule-based approaches

In many scientific disciplines, human knowledge is systematized in relations of variables and represented by mathematically formalized models. Examples include equations of control theory in robotics or models for calculating ocean currents in climate research. Knowledge can be represented by if-then rules in less complex contexts or as heuristics for search procedures and by sets of rules in formal logic, taxonomies, ontologies or knowledge graphs – such as the systematized representation of relationships between places, entities (e.g. people, companies) and events. Such approaches process knowledge automatically in some cases. However, they do not draw any (independent) conclusions (reasoning).

Knowledge-based AI (symbolic AI)

The aim of knowledge representation and processing is to represent existing knowledge using symbols (e.g. words, signs) in such a way that computers can process it. They perform tasks or generate new knowledge by sorting, searching or linking symbolic information and drawing their own logical conclusions. Examples include methods for the automatic creation of logic-based programs such as Inductive Logic Programming (ILP). Symbolic AI was the leading AI approach in the 1960s and 1970s, but then reached its limits (table 1).

Table 1: Advantages and disadvantages of symbolic AI

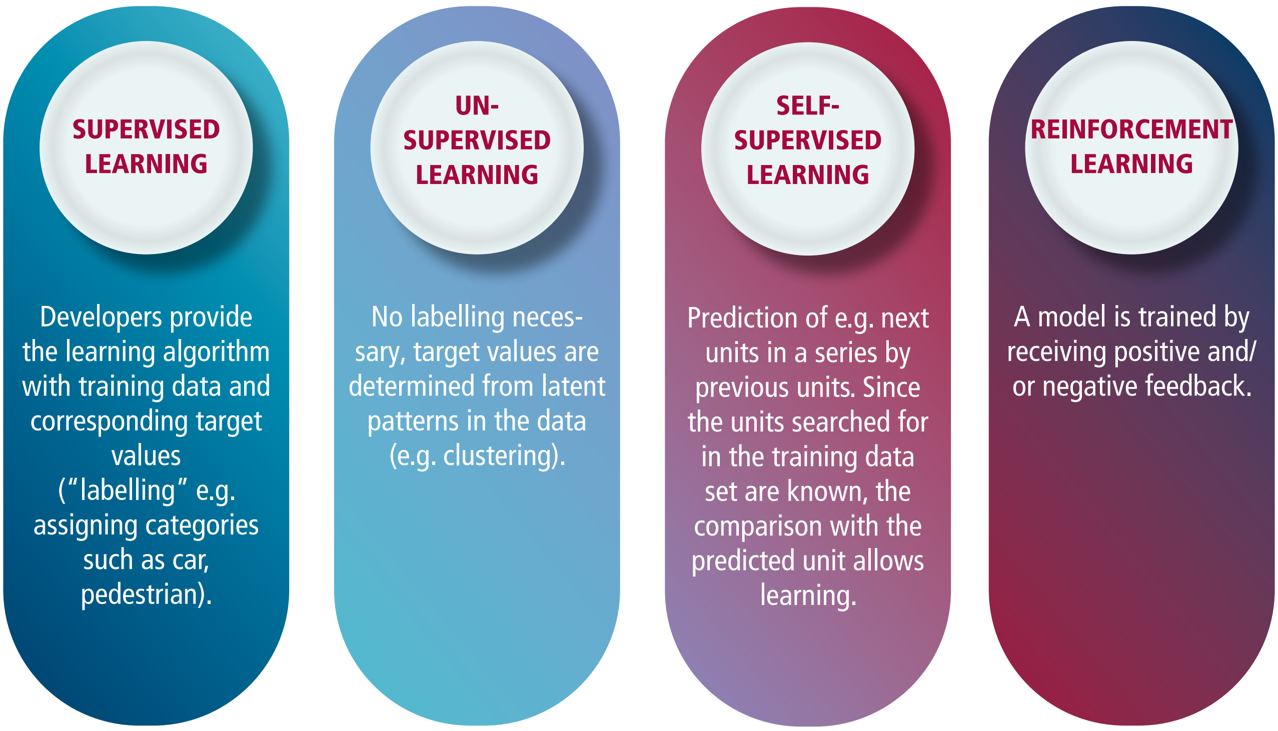
Advantages	Disadvantages
Existing human knowledge can be incorporated into the development and operation of systems	**Cost-intensive**, as a lot of manual coding is required to convert problems into input for the systems
Transfer of knowledge units possible for similar applications	**Implicit knowledge difficult to represent**; not all conceivable cases can be implemented in advance
Results are **explainable**, conclusions are **comprehensible**	**Difficult to scale**; difficult to generalize for different applications
Process can be checked and errors can be corrected and formal correctness verified	Poor **performance** with large empirical data streams (in real time)
Best **suited for well-definable and static problems** as well as for processing and modelling abstractions	Systems are poor at dealing with new and **contradictory situations**
	Maintenance may be difficult, as complex verification and validation required

Own compilation based on:
Ilkou & Koutraki (2020); Saker et al. (2021).

Starting point for AI systems: Collected data

One contrast to symbolic AI is data-intensive Machine Learning, which belongs to the sub-symbolic AI paradigm. This involves learning statistical models based on quantitative correlations in large amounts of data in order to perform certain tasks (e.g. making distinctions). The aim is to map input data to output data or target variables through mathematical formalization via a function, e.g. in the case of images from road traffic to categories such as "cars" or "pedestrians". The model is learned from the data and not created by humans.

For a long time, less complex types of Machine Learning were used, such as Support Vector Machines. Currently, artificial neural networks dominate, which are modeled very abstractly on the function of biological neurons. If the artificial neurons are activated during training, this strengthens certain connections in the network. The information is therefore represented by the different strengths of the connections, also known as weighting. In this way, a model is trained that can map the relationships between input data (e.g. images) and output data (e.g. classifications). If there are several network layers between the input and output layers, this is referred to as Deep Learning. Different learning methods are used within data-driven AI:



While considerable success has been achieved in the last decade with Supervised Learning in particular, large neural networks generated by self-supervised learning impress with their performance now – for example, language models that generate texts according to a prompt. This trend also shows that AI models can no longer only perform specific tasks, but a wide range of tasks (e.g. creating essays, program code or translations). However, challenges in terms of explainability and logical or mathematical capabilities have not been solved (Löser & Tresp et al., 2023).

Table 2: Advantages and disadvantages of data-driven AI

Advantages	Disadvantages
Very suitable for **solving perception and control problems**	**Resource and computationally** intensive: existing human knowledge must be relearned
Implicit regularities and correlations are learned from data, so **less knowledge is required front up**	Required large data volumes and **high data quality** are often not available
High **scalability**, especially with newer self-supervised learning processes	Data selecting and pre-processing requires **explicit human knowledge** (e.g. medicine)
High **performance**	Poor **generalizability** beyond the feature distribution in the training data set
Robustness in the event of incorrect, missing or insignificant data	**Biases** in training data sets can be reproduced (prejudices, discrimination, etc.)
Reusability of large models (e.g. language models) and model customization for specific tasks and domains possible	Poor **interpretability or explainability**: so-called black box models, problematic especially in safety-critical areas (e.g. medicine, mobility)
	Weaknesses in **logical reasoning**
	Data labeling often requires a significant amount of human effort

Own compilation based on:
Ilkou & Koutraki (2020); Saker et al. (2021).

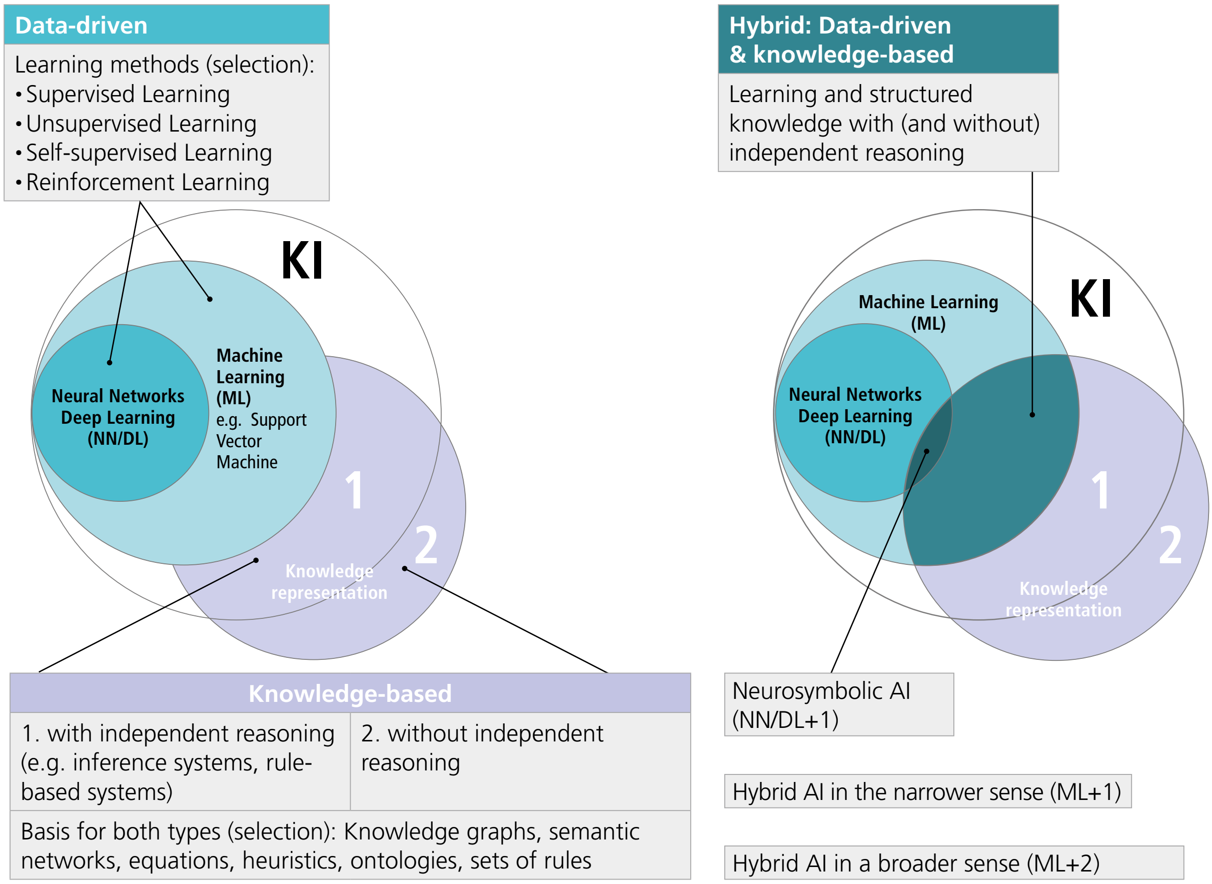
Hybrid AI: Knowledge and data as a starting point

Hybrid AI uses both human knowledge and collected data to develop solutions to problems. The following approaches are possible:

- **Combination of knowledge-based approaches with Machine Learning**, e.g. physics-informed Machine Learning (hybrid AI in the broader sense).
- **Combination of different paradigms of AI research**, i.e. symbolic and sub-symbolic AI (hybrid AI in the narrower sense).

Both approaches have advantages and disadvantages and can benefit from each other (table 1, 2, 3). With the recent successes of Deep Learning, research into its combination with traditional symbolic AI has grown in particular. Hybrid AI systems are therefore currently also being discussed under the term neurosymbolic AI (figure 1). They **promise the best of both worlds**: Ideally, neural networks retain their trainability and effectiveness with partially erroneous data sets, while knowledge-based components contribute explainability and the simple integration of human knowledge.

Figure 1: AI paradigms, AI types and learning methods



Source: Own compilation based on Goodfellow (2016). The interaction of Machine Learning (ML), structured knowledge and reasoning can be quite complex for hybrid AI in the respective application context, see van Bekkum et al. (2021, p. 6 ff.). In a robotic system, for example, ML, rules for behavioral constraints (see 2) and reasoning based on rule-based relationships between ML output and further output of analysis (see 1) can interact (figure 4).

Table 3: Examples of hybrid AI

Application	Knowledge-based AI	Data-driven AI
AlphaGo (2016) The AI system beats the best players of the Chinese game "Go"	Monte Carlo search in decision trees	Supervised and Reinforcement Learning
Google (as at: 2020) Pragmatic approach to optimizing internet searches	Systems for manipulating symbols, e.g. tools for navigating in knowledge graphs	a.o. BERT
AlphaFold2 (2020) Prediction of the 3D structure of proteins	Symbolic forms of representation of the physical 3D structure of molecules	Transformer
Patel & Ott (2022) Prediction of tipping points that dramatically change complex systems (e.g. weather, climate)	Knowledge-based modelling of physical laws	Reservoir Computing
Kirchner et al. (2019) Basic functions of a robot are modelled with control technology, while perception and learning ability are based on ML	Rules and/or parametric equations of control engineering	a.o. Reinforcement Learning, Deep Learning

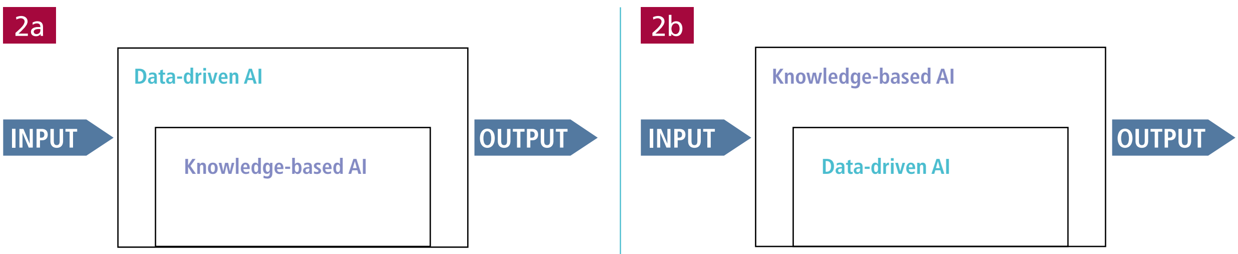
Own compilation based on: Kirchner et al. (2019); Marcus (2022); Patel & Ott (2022).

Exemplary types of hybrid AI

Of the various possibilities for hybrid AI systems, two types of combinations are presented in simplified form below (for more comprehensive typologies, see Kautz, 2021; Ilkou & Koutraki, 2020; Saker et al, 2021; Hamilton et al., 2022, and for more complex combinations see Bekkum et al., 2021).

Nested systems: An AI system consists of a data-driven component in which a knowledge-based system is embedded – or vice versa (figure 2a, b).

Figure 2: Nested Systems

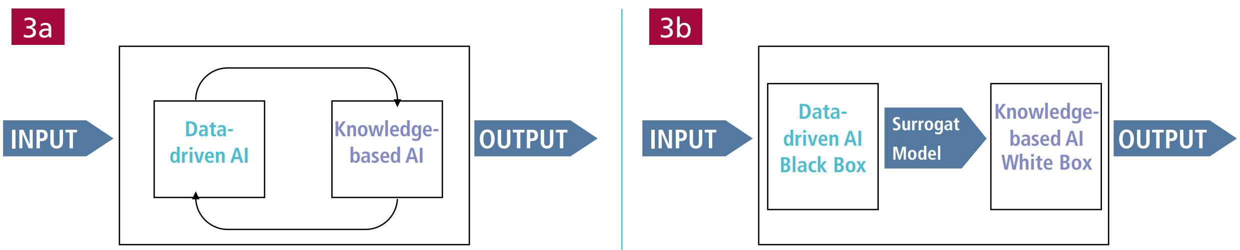


Example: Based on learned relationships from knowledge graphs Graph Neural Networks (GNN) draw conclusions by predicting relationships (e.g. risk management for supply chains).

Examples: Within a heuristic search in a decision tree a neural network is used for pattern recognition (e.g. AlphaGo), applications in robotics and autonomous driving.

Cooperative systems: Input is fed into a knowledge-based system, the output of which in turn provides input for a data-driven system – or vice versa (figure 3a). The result is an AI system in the form of a mutually informing cycle. Semi-cooperative arrangements are also possible (figure 3b).

Figure 3: Cooperative and semi-cooperative systems



Example: The results from knowledge-based approaches inform Machine Learning, the results of which inform or complement or adapt the knowledge-based component.

Example: AI model to increase the explainability of a data-driven black box model using a symbolic surrogate Model. The latter can approximate the entire black box model and explain its predictions and results globally (e.g. medical diagnostics, figure 5).

Integrate human knowledge interactively: Application examples

Hybrid AI for more explainability

Among other things, hybrid approaches make it possible to combine implicit knowledge for classifying information with explicit knowledge from complex, relational contexts. At the same time, the learned symbolic rules can be used to explain neural networks and thus make it possible to understand, which information lead the black box model to certain classification. Symbolic models in the form of rules can themselves serve as explanations if they are not too complex and the addressees have the relevant knowledge. Otherwise, explanations in natural language form can be generated from the rules. In application areas such as medicine, it may be necessary for learned models to be correctable through human feedback (explainable interactive learning). This requires the comprehensibility of model decisions.

Figure 4: Improved explainability in medical diagnostics

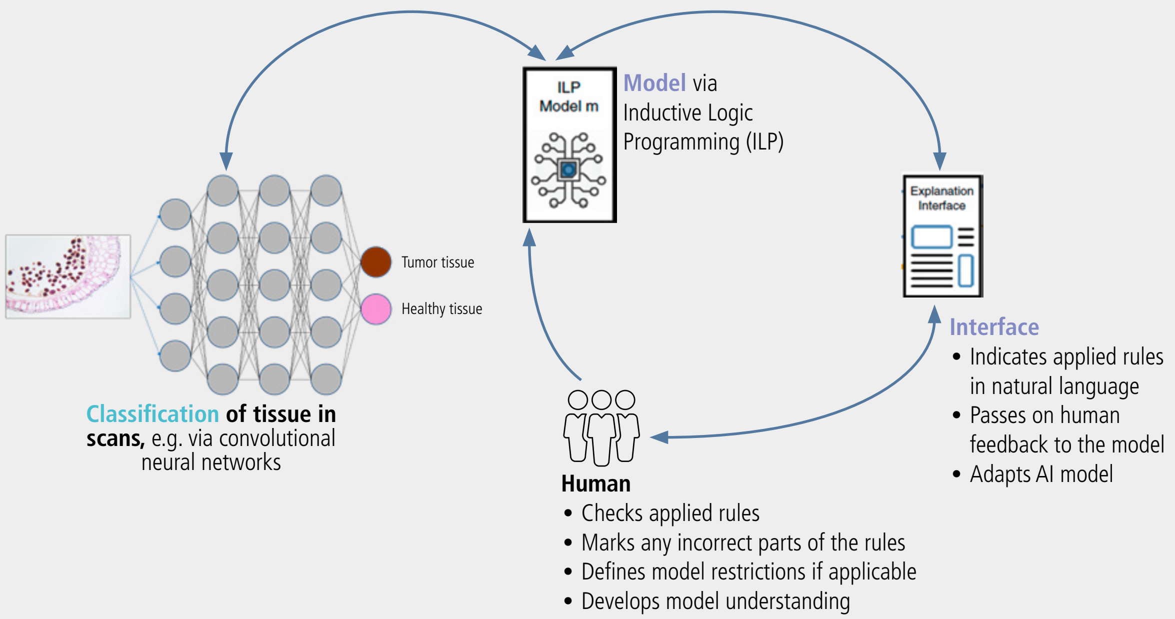
Task
- Support pathologists in the diagnosis of tumors
- Recognize tumor tissue and put it into context

Goal
- Making AI decisions transparent and understandable
- Improve model

Solution
- Tissue scans are **classified with data-driven AI**
- **Rule-based AI** (via ILP) creates comprehensible and correctable AI model
- **Graphical user interface** for interaction between AI model and pathologists

Detection and explanation of tumors through human-AI partnership

Execution: By contrastive comparison of the classification results, a model is learned which, based on background knowledge, learns rules regarding the spatial arrangement of tissue in the scans.



Source: Own compilation based on Schmid and Finzel (2020).
Turquoise stands for the Machine Learning component and **purple** for the knowledge-based component.

Embodied hybrid AI in the service of humans

Robotic systems, understood as embodied AI, have a physical presence, are directly connected to their environment via sensors and can gather information from it. Their basic functions are often guaranteed by rule-based models that are based on human knowledge. Examples of this are reflexes to overcome obstacles when walking or the restriction of a robot's action corridor by rule sets.

The latter safeguards embodied AI so that it can be used in high-risk areas. The movement of a complex robot is also often made possible by methods from control engineering (e.g. optimal control). Here, models are not learned, but created using equations designed and parameterized by humans. With reinforcement learning, however, such models or rules can also be learned. Finally, basic functions can also be combined with Machine Learning (e.g. for object recognition). Feedback from the technical systems for basic functions can in turn be used to inherently train data-driven AI. It can serve as a kind of "corset" or "corrective" for Machine Learning.

Figure 5: Support for patients after a stroke

Task
- Rehabilitation of stroke patients with signs of paralysis

Goal
- Rehabilitating the brain: Support patients in intentional movement and thus positively reinforce involved brain processes

Solution
- **Exoskeleton enables** patients to exercise independently and sustainably
- **Exoskeleton learns** to interpret human intentions and adapt through interaction with humans
- **Symbolic rules** ensure safe interaction

Human-robot interaction in rehabilitation

Execution: The paralyzed arm is guided automatically without a trainer.
⟶ Gravitational compensation of the weight of the exoskeleton and, if necessary, the arm

Teach-in: Training of movements by therapists

Mirror Mode: Healthy arm guides paralyzed arm over the exoskeleton

Learning: Algorithm learns a model from EEG and muscle information to recognize intentional Movements

Correction: Counteract incorrect movement predictions by embedding them in a system of rules (e.g. rules on gaze direction and arm movement)

Reliability: Logical interference from the temporal sequence of movement planning in the brain, muscle activity and direction of gaze to the movement target

Model-update: Feeding knowledge into Machine Learning through positive or negative feedback

Source: Own compilation based on Kirchner et al, 2013.
Turqouise stands for the proportion of Machine Learning and **purple** for the proportion of knowledge-based components.

Potentials and challenges

Hybrid AI promises more efficient, more robust, more explainable and more trustworthy AI systems that are subject to less bias and require less data for the learning process. The following section takes a closer look at the potentials and challenges:

Table 4: Potentials and challenges of hybrid AI systems

Potentials	Challenges
Performance and transparency: • If rules do not have to be learned in the first place, this can increase performance • Learning with less data; improved coherence and consistency of learned models • Handling scenarios that deviate significantly from the training data • Improved correctability of incorrect model output. Errors are easily recognizable; e.g. for domain experts; fast and uncomplicated model adaptation by feeding in human corrections is possible • Interpretability for users through transparent AI decisions • Improved anchoring of the meaning of words in reality (see grounding problem with Large Language Models)	General: • Hybrid AI systems often not transferable to other areas of application • Lack of guidelines for combining different AI approaches • No consensus on theories of logical reasoning, as with knowledge-based systems • Lack of benchmarks in the research and development of hybrid AI systems
Resource efficiency: • Existing explicit knowledge does not have to be learned in a resource-intensive way	In practice: • The creation of hybrid systems is complex and still associated with open questions (see examples below) • Interdisciplinary collaboration between researchers and domain experts necessary • Cross-domain generalization and standardization of technical solutions
Security: • Use formalizable rules as constraints for (semi-) autonomous AI systems to regulate undesired system behavior	
In practice: • Expectation (Gartner 2022): Widespread use of combined AI approaches including hybrid AI in the medium term, as they are usually cheaper for companies than purely ML-based solution • Example of industrial quality control: pattern recognition using Deep Learning combined with knowledge-based components, to understand the causes of quality deficits	

Sources: Own compilation based on:
Ilkou & Koutraki (2020); Kirchner et al. (2021);
Saker et al. (2021); Beyer & Müller-Quade (2022);
Schmid (2022); Löser & Tresp (2023).

Outstanding issues?

- How can neural and symbolic architectures be effectively integrated, especially to solve complex problems?
- How can symbolic AI problems be solved with Deep Learning?
- How can Deep Learning be improved with knowledge bases, graphs and expressive meta data?
- How can interactive learning with complex symbolic explanations be implemented for an effective human-AI partnership?

Expertise from Plattform Lernende Systeme

While Machine Learning approaches are purely data-driven, hybrid AI allows existing knowledge to be taken into account when learning. In principle, this corresponds to human learning processes: What you already know, you don't have to keep relearning. Hybrid AI leads to greater data parsimony and more robust models.

Prof. Dr. Ute Schmid, University of Bamberg

People are capable of learning. However, they usually act on the basis of what they have learned and reflexes. Accordingly, we should take an "integrative" approach to the development of robotic systems by embedding hybrid AI approaches in rule-based basic functions and function-defining structures. This saves resources and increases the safety of the systems.

Prof. Dr. Elsa Kirchner, University of Duisburg-Essen, DFKI

Further reading

Görz, G., Schmid, U., Braun, T. (2021): Handbuch der Künstlichen Intelligenz (Aufl. 6). De Gruyter. Berlin, Boston. https://doi.org/10.1515/9783110659948

Hitzler, P., Eberhart, A., Ebrahimi, M., Saker, K. & Zhou, L. (06 2022): Neuro-symbolic approaches in artificial intelligence. National Science Review, 9(6). Online at: https://academic.oup.com/nsr/article/9/6/nwac035/6542460

Ilkou, E. & Koutraki, M. (2020): Symbolic Vs Sub-symbolic AI Methods: Friends or Enemies? CIKM. Online at: http://ceur-ws.org/Vol-2699/paper06.pdf

Saker, K. et al. (2021): Neuro-Symbolic Artificial Intelligence. Current Trends. Online at: https://arxiv.org/pdf/2105.05330.pdf

Further sources

Imprint
Expertise: Elsa Kirchner, Ute Schmid
Editors: Max Hösl, Birgit Obermeier
Editorial staff: Lernende Systeme – Germany's Platform for Artificial Intelligence | Managing Office | c/o acatech | Karolinenplatz 4 | D-80333 Munich
kontakt@plattform-lernende-systeme.de | www.plattform-lernende-systeme.de
Status: December 2023 | Photo credits: J. Schabel, UDE - Frank Preuß / S. 10
Follow us on X and LinkedIn.

SPONSORED BY THE

Federal Ministry of Education and Research

acatech
NATIONAL ACADEMY OF SCIENCE AND ENGINEERING