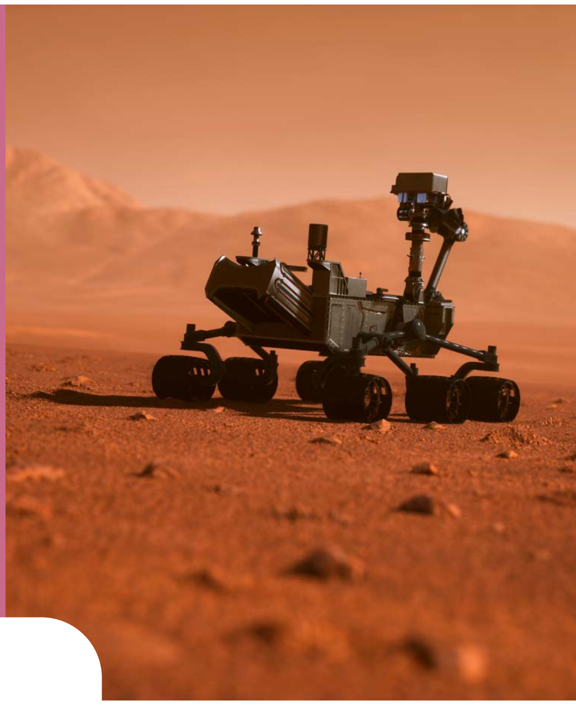
Lernende Systeme

GERMANY'S PLATFORM FOR ARTIFICIAL INTELLIGENCE

# Competent in use
## Variable autonomy of self-learning systems in hostile-to-life environments



White Paper by Jürgen Beyerer et al.
Working Group Hostile-to-Life Environments

## Executive Summary

Whether in space, in the deep sea or in disaster areas - operations in such hostile-to-life environments represent a major challenge with considerable risks for humans. Self-learning systems can help to reduce hazards and risks for humans or to make such environments accessible in the first place. Mission configurations, the degree of autonomy of self-learning systems and the intensity of interaction with humans can vary considerably. A good division of labor is critical to the success of collaboration between humans and self-learning systems. The working group Hostile-to-Life Environments of the Plattform Lernende Systeme has identified and investigated key requirements for this division of labor between humans and self-learning systems as well as for the competence of the self-learning systems in the respective situational context of application.

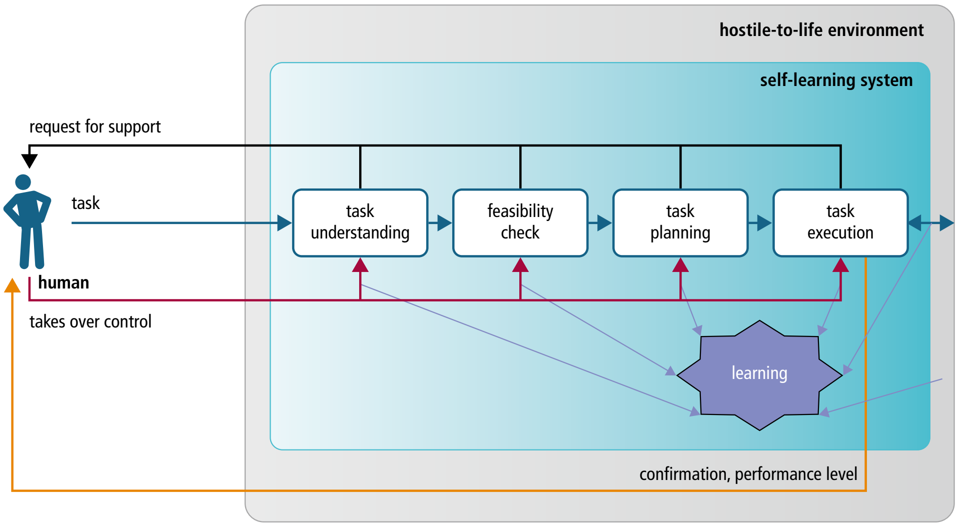
Self-learning systems in hostile-to-life environments have unique properties compared to application domains such as industry and transportation. At this stage, such deployments cannot be designed without humans as the overseers of what is happening. Instead, the goal is to support humans and to minimize their risk of danger. The core idea: as much autonomy as possible – only as much human intervention as necessary.

With reference to practice-oriented use cases, the white paper addresses the questions of why variable autonomy should be aimed for self-learning systems in hostile-to-life environments, which architectural components such systems require, which research needs exist and which issues arise for applications.

## Fundamentals for the cooperation: human and self-learning systems

First, the specific requirements for the degree of autonomy of self-learning systems are defined for their use in hostile-to-life environments. This degree of autonomy depends, among other things, on the environment and the nature of the task, the capability of the self-learning system as well as legal and ethical guidelines for its use. Second, autonomy must be adjusted so that humans only need to intervene where it is necessary and reasonable. In addition, due to the dynamics of deployments in hostile-to-life environments, the systems must be able to adapt their degree of autonomy to the respective situation by themselves, on the one hand, or be adapted by humans, if necessary, on the other hand; this requires highlighting the perspective of variable autonomy during the deployment time of a self-learning system in contrast to previous conceptions of autonomy. The self-learning system is always to be understood as a human-machine system, that is to say, there is an interaction with the human, which can be very direct (e.g. through remote control) or only selective (e.g. through the adaptation of missions).

**A simple workflow scheme for autonomy (an autonomous self-learning system) with varying autonomy levels**




© Plattform Lernende Systeme

Autonomy, control and learning in robotic systems are mentioned as further important foundations for the use of self-learning systems, which concern, for example, basic control modes for robots as well as prerequisites for higher degrees of autonomy of robotic systems. Crucial components of autonomy in self-learning systems are the knowledge of the system's own capabilities and the evaluation of its own actions in relation to the task, the environment, and the current situation. The greatest challenge for a self-learning system – which is supposed to act completely autonomously – is to know all these parameters dynamically over the period of the respective mission and, in addition, to have always the possibility to react in time to new events and to adapt further actions accordingly.

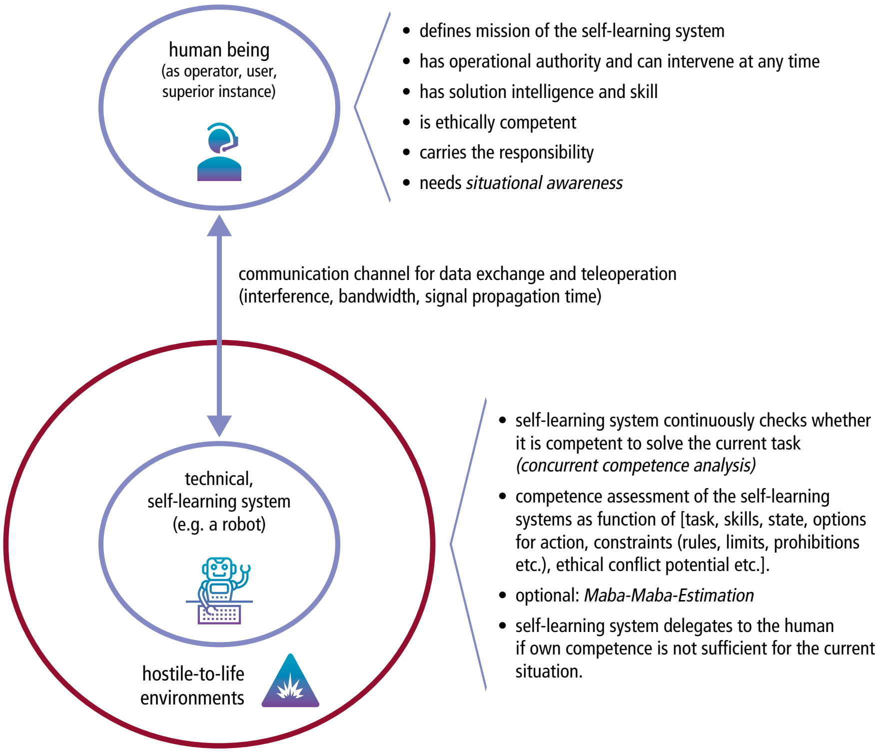
## Variable autonomy levels of self-learning systems

An overview of existing taxonomies for levels of autonomy in conventional domains of application such as medicine and autonomous driving makes clear that autonomous systems in hostile-to-life environments are confronted with special conditions and thus with quite different requirements for autonomy. Existing taxonomies for autonomy in conventional domains usually range from fully (remotely) controlled (non-autonomous) to fully autonomous. Hostile-to-life environments, however, are very diverse and missions have a high variability, combined with very different requirements for autonomy in each situation. Therefore, the whitepaper focuses on variable autonomy; a perspective that goes beyond the previously existing distinction of discrete levels of autonomy. To demonstrate this perspective of continuous degrees of autonomy during ongoing missions, the authors design an architectural model that defines components of an autonomous system and illustrates how different degrees of autonomy can be concretely activated in a variable mode.

## Technical and ethical trustworthiness of self-learning systems

In principle, self-learning systems should be developed in an ethically trustworthy manner regarding criteria such as comprehensibility, security, and the avoidance of non-intended consequences. In hostile-to-life environments self-learning systems may also encounter ethically or legally problematic situations during a mission, which would subsequently also increase the criticality of the self-learning systems in its application context in such situations. Therefore, these systems require, among other things, components that can recognize and communicate such situations. The competence of a self-learning system can change during its use, for example because it is confronted with different conditions. For this reason, the experts propose an ongoing competence analysis during missions for the architecture of self-learning systems to be able to determine reasonable and situationally possible degrees of autonomy. In the relationship between competence analysis and degree of autonomy, a lack of competence always means increased human intervention up to teleoperation (= remote control). Up to the full autonomy, different intermediate levels can be distinguished in which the human being gradually hands over control to the self-learning system. Basically, the case can occur that a system masters a task (e.g. exploration of an unknown terrain), but it is stuck. In this case, its options for action would be limited to such an extent that remote control by humans would become necessary.

**Self-learning systems in hostile-to-life environments as human-machine systems**




© Plattform Lernende Systeme

## Recommendations and outlook

To ensure that self-learning systems can support people even more effectively in hostile-to-life environments in the future, the experts conclude by highlighting options that relate primarily to strengthening research and product innovations. They see a need for research and development regarding the various architectural components and the competence analysis of self-learning systems.

## Architectural Components

Indispensable components in the architecture of a human-machine system for effective, efficient, and responsible use of self-learning systems in hostile-to-life environments are:

- Ongoing analysis whether the self-learning system's competence is sufficient to solve a current task.
- Ongoing analysis whether the self-learning system would be placed in an ethically problematic situation or would have to make ethically problematic decisions.
- Ongoing analysis whether the self-learning system gets into a situation that raises legal issues.
- Component for ensuring situational awareness for the human as a prerequisite for allowing human operators to take control at short notice.

## Competence analysis

To enable self-learning systems to perform comprehensive competence analysis as a kind of self-monitoring, several subfields of artificial intelligence (AI) need to be integrated. Quite a few important milestones would have to be reached here, so that the systems acquire this ability robustly. The most important milestones are:

- The self-learning system creates a match between the task and its own capabilities: in the case of complex tasks, this can include further analyses that today's systems do not yet perform in such a comprehensive way, such as
  - analysis of the environment,
  - detection and consideration of contexts,
  - generalization of one's own capabilities to new tasks,
  - decomposition of a complex task into reasonable subtasks that can be solved,
  - efficient distribution of the subtasks among several learning systems.
- Define a technical framework for how self-learning systems...
  - ...can recognize and communicate ethically problematic situations.
  - ...can recognize legally problematic occurrences.
- Ensuring situational awareness for allowing human operators to take control at short notice.

Ultimately, all these overarching milestones address aligned research questions, including those of transparency and explainability of AI, which are important pillars of national and European AI strategies.

For applying companies it is important to create a legal framework that can provide orientation for innovations and products. Under what conditions can self-learning systems be used for operations in hostile-to-life environments? What are and what are they not allowed to do? This kind of guidance also helps in the areas of a system's competence analysis, for example, to determine when this kind of analysis by the system itself becomes necessary in the first place. Not every system will need this capability in the same way. These questions then, in turn, give impetus to research. Ultimately, the certification of self-learning systems also supports the application here, as it also offers the possibility of creating more transparency in the market and setting standards.

**Imprint**
Editor: Lernende Systeme – Germany's Platform for Artificial Intelligence | Managing Office | c/o acatech | Karolinenplatz 4 | D-80333 München | kontakt@plattform-lernende-systeme.de | www.plattform-lernende-systeme.de | Follow us on Twitter: @LernendeSysteme | Status: January 2021 | Photo credit: designprojects/AdobeStock/Title

This abstract was created based on the white paper *Competent in use – variable autonomy of self-learning systems in hostile-to-life environments*, Munich, 2021. It was created by a team of authors from the working group Hostile-to-Life Environments of Plattform Lernende Systeme. The original version of the publication is available online at: https://www.plattform-lernende-systeme.de/publikationen.html

SPONSORED BY THE

Federal Ministry of Education and Research

acatech
NATIONAL ACADEMY OF SCIENCE AND ENGINEERING