Criteria for Human-Machine Interaction with AI

White Paper by Norbert Huchler et al.
Working Group Future of Work
and Human-Machine-Interaction

Executive Summary

The use of Artificial Intelligence (AI) offers a wide range of potentials for safe, autonomous and self-determined work as well as attractive and competitive jobs. For example, AI-based assistance systems can relieve employees of strenuous or dangerous activities and support them in complex processes and decisions. At the same time, AI systems are changing the interaction between people and technology in our working environment. In the future, people and machine will interact even more strongly – and differently – than in the past, since Machine Learning (ML) and similar technologies enable machines to perform certain tasks independently and to learn continuously.

The increasing collaboration between humans and technology makes it necessary to readjust the distribution of work when using Artificial Intelligence. In order to shape this collaboration in the interest of human beings, technology must tie in with the advantages and potentials of human thought and action and place mutual complementation – and not replacement or conflict – at the centre of interaction. A coordinated balance which considers the employees as well as to the technological and economic potentials of Artificial Intelligence increases the chances for an individually and socially accepted use of Artificial Intelligence in the working environment.

Only clearly defined orientation standards for the new distribution of work can secure jobs be created, qualified workers be trained, and the design of good and humane work be implemented. This is where the present catalogue of criteria for human-machine interaction working environment comes into play, which was drawn up by experts from the working group Future of Work and Human-Machine-Interaction of Plattform Lernende Systeme. The criteria aimed at a sustainable future-oriented and human-centred design of human-machine interaction can be summarised in four clusters:

Criteria for the design of human-machine interaction

Protection of the individual	
Safety and health protection	• Avoiding risks to the physical and mental health of employees • Protection against accidents and damage (personal injury and property damage) • Prevention of negative physical or mental stress consequences
Data protection and responsible performance recording	• Protection of personal rights, data economy and earmarking of data use • Avoidance of data analysis for unjustified performance measurement • Development of a positive culture of performance feedback • Transparency of and ability to analyse and use data
Diversity sensitivity and freedom from discrimination	• Protection against discrimination of individuals or groups • Existing legal system as basis for diversity sensitivity and freedom from discrimination
Trustworthiness	
Quality of the available data	• Avoidance of data with insufficient quality • Prevention of distorted data sets, errors/misinterpretations and discrimination • Improvement of the human-machine interaction through reliable data
Transparency, explainability and consistency	• Implementation of approaches of an explainable Artificial Intelligence (Explainable AI) • Development of methods for the traceability of self-learning systems • Creation of (graded) transparency of decision-making processes of self-learning systems • Prevention of demotivation through consistent design of the human-machine interaction
Responsibility, liability and system trust	• Transparency about and attributability of responsibility • Competence and control over the system as a prerequisite for taking responsibility • The level of control of the system as a measure of the nature and extent of responsibility • Starting points along the concepts of a trustworthy Artificial Intelligence (Trust-worthy AI)
Meaningful division of work	
Adequacy, relief and support	• Appropriate work content and requirements • Complementary completion of human and machine capabilities for a meaningful division of work • Sustainable relief and support of employees through AI-based assistance systems • Ability to work with AI systems
Ownership of the company and situation control	• Targeted and transparent design of action and situation control • Minimization and prevention of risks and negative consequences of stress
Adaptivity, fault tolerance and customizability	• Enabling self-learning systems to adapt flexibly and situationally to the needs and requirements and to the working practice of the users

Convenient working conditions	
Scopes for action and substantial work	• Safeguarding and, where appropriate, scope for action for employees (autonomy and freedom of decision as well as diversity of action) • Inclusion of basic human needs for meaningful, motivating and health- and personality-promoting work
Promoting earning and experience	• Enabling the mutual learning of humans from the machine and vice versa • Comprehensible, adaptive design of systems for the integration of knowledge and experience • Ensuring the transfer of data into information or from information into knowledge
Communication, cooperation and social bonding	• Double sensitization of Artificial Intelligence to social contexts and structures • Support of interpersonal communication, cooperation and solidarity • Artificial Intelligence as a cooperation partner

These criteria are addressed to actors involved in the planning and development of self-learning systems and at actors involved in the implementation of AI systems in companies. The catalogue of criteria is intended to provide guidance for the design of the division of work between humans and technology in the application of self-learning systems. In addition, the criteria are intended to inspire the further development of existing regulations – for example in standardisation, legislation or social partnership – and should enable more flexible, self-determined and autonomous work in future.

SPONSORED BY THE

Federal Ministry of Education and Research

acatech

NATIONAL ACADEMY OF SCIENCE AND ENGINEERING