



## Discussion paper for the expert dialogue “HTI – Working with Artificial Intelligence”

# “Democratic Technology Design in the Digital Transformation”

### Abstract:

As part of the expert dialogue “Human-Technology Interaction – Working with AI” initiated by the Observatory for Artificial Intelligence in Work and Society (AI Observatory), stakeholders from science, trade unions, companies, civil society and associations work together to develop joint strategies for a good human-technology relationship for the use of artificial intelligence (AI) in work processes. The discussion paper describes the problem and scope of the expert dialogue while at the same time contributing to the debate surrounding the human-centric design of AI. The discussion paper is the result of cooperation coordinated by the BMAS between the Institute for Innovation and Technology (iit) and the participants of the expert dialogue.

### Technology – work – society: the conflicting priorities of opportunity and risk

Since the onset of industrialization, technology has played an integral role in the world of work. With the advent of technology following the development of domestic and home electronics, technology has also increasingly shaped the personal and social lives of people since the 1960s and 1970s and has now entered their everyday lives and world of work to a significant extent. Due to their efficiency and complexity, digital technologies and tools are associated with positive expectations of further improvements to working and living conditions as well as with reservations and uncertainties. This ambivalence in the individual's and society's relationship with technology is leading to growing requirements regarding these new tools, for instance, in terms of security, transparency and control by users. How an enlightened and autonomous interaction by humans with AI in the future can be achieved in practice is a key issue. It also determines how inclusive, open and democratic digital work and society will become in the future.

Against this background, there are some fundamental questions that must come to the fore in the social discourse: how are digital technologies changing our society? What possibilities and instruments are available to society to design these technologies? Which new practices for the introduction of and interaction with digital technologies are required? How can we reach a broad consensus regarding the goals, possibilities and instruments to design technology in a future-proof manner? The question also arises in this context as to how digital technologies and applications based on them can be reconciled with our democratic self-perception and designed with the established decision-making processes.

### AI opens up new opportunities and raises new questions

The questions formulated above arise in particular in conjunction with technologies that fall under the heading of “artificial intelligence”. A large number of future innovations and applications that influence our everyday and working lives will be based on AI. In this context, the application of AI differs in important respects from previous technological developments.

Artificial intelligence enables the use of large quantities of data that humans themselves can only evaluate with great difficulty or not at all. Furthermore, AI-based systems can perform simple, standardized and repetitive tasks, for example, faster, more cost-efficiently and more reliably than humans. As their development progresses, AI-based systems will be able to take on an increasing number of cognitive tasks that have been reserved for humans up to now.



On the one hand, humans and AI-based applications will enter a competitive relationship with one another in some cases. On the other hand, this will create new possible ways for technologies to provide humans not only with physical assistance but increasingly with cognitive assistance. In addition, “hybrid systems” will emerge in which technology will now also make decisions alongside humans.

Therefore, the use of AI-based systems fundamentally changes the relationship between humans and machines and creates new socio-technical connections. The use of AI-based systems raises complex ethical, social and legal issues concerning the design of human-technology interaction all the way to fundamental questions of sovereignty and autonomy.

## How human should AI be?

The possibility of guaranteeing final decisions by a human is one of the central requirements raised for the future interaction with AI (German Confederation of Trade Unions (DGB) 2020). This applies above all for decisions that have consequences for humans, for example, regarding the distribution of work tasks. In this case, a significant challenge arises in terms of the direct interaction (interface) between humans and AI-systems themselves, whose design largely determines the use and impact of the technology.

Therefore, a particular focus is on the interaction itself, through which human wishes, feelings and needs are projected in the form of technical systems. This can cause people to interact with AI systems as if they were social stakeholders. Even if developers do not intend an AI to send out social signals, users can still project these signals into it (Bovenschulte 2019). The anthropomorphism of artificial intelligence can cause people to transfer social norms and values to their technical interaction partners, and to assume intentionality – thereby becoming susceptible to misplaced trust in an AI, which can extend all the way to possible manipulation by AI (Bartneck et al. 2019). Therefore, the premise that AI systems are man-made artefacts like all technologies is of fundamental importance for the aspiration for human-centric technology design of AI. In other words, humans (developers, decision-makers in companies as well as users) decide on the design and use of AI and therefore have a decisive influence on what the working society of tomorrow will look like. For this reason, the design of AI, as with other future technologies, is a long-term challenge for society. It is not only the technology that is designed in this case. The use of technology, the work organisation and the role of humans in the work process are also being designed.

## How can aspiration become reality?

To harness the design opportunities offered by the use of AI, AI systems should be designed to be “responsible” and “human-centric”. The Federal Government of Germany is also committed to this ambition with its AI strategy (Federal Government 2018; German Bundestag 2020). But what does this mean specifically, and how can this aspiration become reality?

Today, organisations from the private and public sectors have already formulated numerous principles, guidelines and catalogues of rules for the development, introduction and application of AI systems (see among others DIN/DKE 2020; Ethics Committee HR Tech 2020; Huchler et al. 2020; Puntschuh/Fetic 2020). The vast majority of these adhere to the principle derived from a more recent understanding of technology design whereby users are incorporated at an early stage and systems are designed with a focus on the (desired) solution to a problem rather than simply on what is technically feasible. The main argument here is that the technology should serve humans.

Nevertheless, the conventional understanding of technology design whereby users are only considered after the technology has been decided on, that is, when it is actually rolled out in a company, is often still the dominant business rationale. There are many possible reasons for this. The basic principles of human-centric AI, such as reproducibility, fairness and the protection of personal data, as well as specific requirements such as human oversight and ultimate decision-making authority can create a conflict of objectives with regard to the economic potential of AI systems.



This applies in particular given that stakeholders who develop and use AI-based systems are often competing on the global stage. The rapid pace technology development and the slow process of societal negotiation regarding the rules and boundaries for the design and use of AI systems can further intensify this conflict of objectives. For this reason, the expert dialogue addresses, among other things, the goal of debating how aspiration can become reality and how human-centric technology development can be reinforced in practice.

## **Change must take place in the system**

Meeting the aspiration of human-centric AI requires more than policies and prescriptive catalogues of rules for AI development. On the contrary, design principles for human-centric AI development must permeate all aspects of technology development and must be brought to life through practical actions by the stakeholder groups in society and in the world of work. This requires a change in the entire (innovation) system used to develop an AI application. In this context, all stakeholders in the innovation system must measure themselves against the aspirations of human-centricity and the orientation towards the common good. This applies irrespective of whether an AI-based system is used as a support system at the workplace, to decide on credit terms, or to help people overcome the challenges of everyday life.

## **Human-centric technology design as a systemic task: the expert dialogue “Working with AI”**

With the expert dialogue “Working with AI – Social and Technology Design”, the Observatory for Artificial Intelligence in Work and Society (AI-O) is addressing the challenges outlined above. In doing so, it makes reference to the established dialogue between stakeholders from science, technology development, companies and trade unions as well as from civil society. The expert dialogue shifts the focus to the question of how principles of human-centric technology design can be developed and transferred from theory into practice. What options and instruments for designing digital technologies are available to society? And what new methods of development, introduction and application are required? What systemic changes in the innovation system are required to meet the aspiration for human-centric AI? And what new requirements do digital technologies and the applications and tools that they underpin place on the design of technology and its use in companies in the context of the social partnership?

The expert dialogue will take place during the first half of 2021 in the form of digital workshops. In preparation for this, the participants together with the BMAS and the Institute for Innovation and Technology (iit) have identified topics that will be processed in a structured manner over six events.

### **1. Human enhancement**

The debate surrounding AI and its impact on the world of work focuses in particular on increases in efficiency and effectiveness and how these affect companies and employees. The consequences for employees in this context are discussed mainly in terms of possible negative effects, such as the devaluation of skills, substitution effects, increased workloads, as well as the use of technical systems for the purpose of monitoring and controlling performance. While the awareness of possible dangers to the world of work from the introduction of AI is important and necessary, the purpose of the “Human Enhancement” workshop is to help identify the positive potential of artificial intelligence. The central question is how human-AI interactions can be designed in a manner that reconciles the interests of business and of employees in order to achieve maximum benefits for society. The starting point is the investigation of the possibilities where AI can enhance the physical and cognitive – or knowledge-based and decision-based – abilities of humans. In this context, the extension of human possibilities for action (expanding the spectrum of physical and cognitive actions) as well as the possibility of experiencing workplaces enhanced by AI in a positive light (e.g. through design that enables the experience of self-efficacy) at individual level and collective level will be discussed (including the performance capabilities of human-machine teams in relation to economic and social goals; inclusion).



## 2. Culture as a Tool & Testing Ground for Social Technology Development

Art and culture reflect and use the new digital environment in an experimental manner long before it becomes part of the public discourse on technology. Today, creative artists are already using AI systems to create, share and market art and culture. Consequently, the cultural sector can be regarded as a testing ground for social technology design. In terms of how it is perceived, culture is an important tool for promoting society's adoption of technology. In particular, the film and video-game industries have featured the phenomenon of intelligent machines as themes for decades. By doing so, they have contributed significantly to society's perception and awareness of AI.

Based on the far-reaching disruption that AI systems have already caused in the music and film industry, the workshop "Culture as a Tool & Testing Ground for Social Technology Development" will discuss how the experience of this field of application can serve as a starting point for a critical, design-oriented debate about the design of human-AI interaction in the world of work. In addition, the workshop participants will highlight the contribution of the cultural sector to the process of adoption of technology in society and how it can create a resonance chamber for a progressive, design-driven debate about technology.

## 3. Digital Literacy

Digital systems, technologies and media have entered virtually every area of our everyday private and professional lives. The application of knowledge and the competent use of these systems constitute a new cultural technology. When it comes to designing human-AI interactions, there is a need for current and systematic approaches to define modern digital literacy as well as role-dependent skills profiles and training approaches. In particular, this raises the question as to which skills in working with AI employees must acquire and how they should acquire them so that they can engage autonomously with new technologies. The "Digital Literacy" workshop will identify skills profiles for specific roles (e.g. AI developers, AI users) and derive specific skills requirements. In an additional step, the workshop attendees will discuss which forms of skills acquisition are expedient for the respective target groups based on their specific prior experience and the skills profiles they need to achieve. The workshop will also discuss the extent to which existing internal and external further training instruments require further development.

## 4. Social cohesion for digital work and society

From a social perspective, it is important for as many sections of society as possible to avail of the opportunities presented by digital technologies. Accordingly, it is also important to achieve high and inclusive participation in the world of work. Different dimensions of participation are relevant in this case. In addition to opportunities to acquire skills to develop digital literacy and the granting of fair access to digital assistance systems while preserving digital sovereignty, the participation of employees in the design of the working process and the protection of their freedom to take action and make decisions are also important. The workshop "Social Cohesion for Digital Work and Society" will investigate how companies and employee representatives can help guarantee that all their employees can avail themselves of digital resources irrespective of their duties and roles within the company. Furthermore, the extent to which digital participation can contribute to social cohesion and how technologies can be designed to support social cohesion in a work context will be discussed. The discussion will highlight which forms of work organisation contribute to social cohesion in a working environment characterised by AI.

## 5. Human-centric innovation system

Promoting and enlarging the network of public and private institutions and companies that initiate innovations through a process of interaction is of central importance for human-centric technology development. The innovation system must be configured to take the entire socio-technological system into account when new technologies such as weak AI are developed and introduced. Specifically for technologies that have disruptive impacts on broad sections of society, iterative exchange processes between science/theory and the implementation practise in the relevant field of application are necessary. The aim of the workshop "Human-Centric Innovation System"



is to come up with proposals as to how this iteration process can be designed for the development of AI. The workshop will investigate how the practise level in companies can be motivated to actively introduce the iterative transformation process as part of a co-design process with the employees involved. This calls for closer cooperation between research and companies. These areas must be inspired to enter into a participatory design process. In this case, it is important to describe requirements for the AI system being developed and to formulate criteria for the subsequent functioning of the social-technological system as part of a responsible research and innovation management strategy.

## **6. Human-centric technology design through the reorientation of company innovation processes**

The discussion about the societal consequences of algorithms or automated decision-making has gained momentum. A range of policies and guidelines for the ethical design of AI has been published by various stakeholders from business, civil society, science and politics. From voluntary commitments by major tech companies to comprehensive technology standards all the way to recommendations by committees convened by governments, these policies and guidelines display a high degree of heterogeneity. They differ significantly in length and degree of detail. However, specific principles such as transparency, protection of privacy or fairness and justice appear in virtually all of these publications. At present, it remains unclear how the principles developed, which – given their fundamental nature – achieve a broad consensus, are to be implemented in reality. The workshop “Human-Centric Technology Design Through the Reorientation of Company Innovation Processes” will discuss which conditions are required to make use of AI in the world of work in a meaningful, value-adding way and to design ‘good work’. In the process, existing AI principles, criteria and process guidelines will be placed in context with their practical implementation. How do companies currently organise innovation processes in the area of AI? How do companies view the AI ethics principles? Where do they see opportunities and obstacles to their implementation? Which measures can support innovation processes in the interests of ‘good work’?

### **Results of the expert dialogue**

The results of the expert dialogue should be documented in the form of a common final publication with articles by experts from BMAS and iit. The results of the expert dialogue are intended to serve as stimuli for companies, trade unions, science, politics and developers of technology as to how AI systems can be designed in the future to do justice to the ambition to achieve trustworthy and human-centric AI.



## Masthead

The expert dialogue “HTI - Working with AI” will be conducted as part of the Observatory for Artificial Intelligence in Work and Society (AI-O) of the Policy Lab Digital Work and Society. The Policy Lab, an agile organisational unit of the Federal Ministry of Labour and Social Affairs (BMAS), bundles projects and processes covering all aspects of the digital transformation within BMAS with the aim of developing a clearer picture of the working society of the future.

www.denkfabrik-bmas.de/en  
www.ki-observatorium.de  
@denkfabrik\_bmas  
#AIObservatory  
#humantechnologyinteraction  
Berlin, 03/03/2021

## Authors

Markus Dicks, Robert Peters, Andrea Altepost, Doris Aschenbrenner, Michael Burmester, Detlef Gerst, Carla Hustedt, Bruno Kramm, Matthias Peissner, Olli Suchy, Martin Westhoven, Carolin Wienrich, Marc Wittlich

## Literature

- Bartneck, C.; Welsh, S.; Lütge, C.; Wagner, A. (2019): Ethik in KI und Robotik.
- Bovenshulte, M. (2019): Digitale Lebensgefährten – der Anthropomorphismus sozialer Beziehungen. TAB (Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag)
- BReg. (Bundesregierung) (2018): Strategie Künstliche Intelligenz der Bundesregierung.
- BT (Deutscher Bundestag) (2020): Strategie Künstliche Intelligenz der Bundesregierung - Fortschreibung 2020. Unterrichtung durch die Bundesregierung. Drucksache Nr. 19/25095
- DGB (2020): Künstliche Intelligenz (KI) für Gute Arbeit. Ein Konzeptpapier des Deutschen Gewerkschaftsbundes zum Einsatz von Künstlicher Intelligenz (KI) in der Arbeitswelt.
- DGB (Deutscher Gewerkschaftsbund) (2016): Wir müssen ein digitales Prekariat verhindern. Plattform-Ökonomie. 27.9.16, <https://www.dgb.de/themen/++co++022acff6-84be-11e6-9ce6-525400e5a74a> (5.8.2019)
- DGB (Deutscher Gewerkschaftsbund) (2019): Künstliche Intelligenz und die Arbeit von morgen.
- DIN; DKE (2020): Ethik und Künstliche Intelligenz. Was können technische Normen und Standards leisten?
- Ethikbeirat HR-Tech (2020): Richtlinien für den vertrauensvollen Einsatz von Künstlicher Intelligenz und weiteren digitalen Technologien in der Personalarbeit.
- Huchler, N.; Adolph, L.; André, E.; Bauer, W.; Bender, N.; Müller, N.; Neuburger, R.; Peissner, M.; Steil, J.; Stowasser, S.; Suchy, O. (2020): Kriterien für die Mensch-Maschine-Interaktion bei KI. Ansätze für die menschengerechte Gestaltung in der Arbeitswelt.
- Puntschuh, M.; Fetic, L. (2020): Praxisleitfaden zu den Algo.Rules. Orientierungshilfen für Entwickler:innen und ihre Führungskräfte. Bertelsmann Stiftung; iRights.Lab