

# **Artificial Intelligence (AI)** **for Good Work**

**March 2020**

# Artificial Intelligence (AI) for Good Work

March 2020

## 1 The objective: implementation of ‘Good Work by design’

In January of 2019, the German Confederation of Trade Unions (DGB) published an initial discussion paper outlining a model for the design of AI applications in the workplace, ‘Artificial Intelligence and the Work of Tomorrow’. ‘Good Work by design’ suggests a forward-looking approach to work design in the context of autonomous software systems: ‘The design must begin in the conception and development phase of AI systems. A prerequisite for good design is a broad participation process, which should begin with the definition of the objectives for the AI and its application and should include an impact assessment.’

This approach was met with a broad response. For example, the ‘Platform Learning Systems’ (Plattform Lernende Systeme) AI platform of the German Federal Government or the ‘HR Tech Ethics Advisory Board’ (Ethikbeirat HR-Tech) agrees with the approach in principle. Similar approaches to process-oriented, preventive AI design have also arisen in a social context (for example the Bertelsmann Foundation ‘Algorules’: rules for the design of algorithmic systems) or as the results of labour research projects (Work Implementation Aids 4.o).

This concept paper aims to elaborate on the DGB’s approach ‘Good Work by design’. The focus will be on questions and sticking points regarding the design of change processes in the context of AI use. Based on the ‘Evaluation List’ for guidelines of the HLEG on artificial intelligence of the EU Commission, which was developed for the general use of AI, the focus should be placed on the implementation and its framework conditions at this point.

The concept paper aims to provide a guiding framework for trade union, political and company discussions. At the same time, it’s aim is also to close the legal gaps in the framework conditions for innovative negotiation processes.

## 2 The state of political debate

The debate on ‘artificial intelligence’ (AI) for the workplace has rapidly gained momentum in the last two years. Even if ‘AI’ is often simply used instead of ‘digitalisation’ in public discourse, great potential is in fact attributed to AI applications – i.e. learning software and decision-making systems. This applies both to automation and to intelligent assistance for human work. The international competitive pressure for the development and application of AI in the economy and labour market has increased noticeably. Nevertheless, many companies, including those in Germany, are still in the early stages of using AI-based possibilities. One reason for this is the high level of uncertainty in dealing with AI applications. At the same time, there are concerns ranging from fear of job loss to loss of control and ‘disenfranchisement’ by AI systems both in society in general and among employees in particular.

Policy approaches have a strong focus on trust and acceptance of AI, but remain generalised so far. At both the international and the European level, the central objective is ‘trustworthy AI’. In May 2019, the OECD agreed on principles to ‘respect human rights and democratic values’. The focus here is on fundamental values such as social wellbeing, sustainability, fairness, security or accountability. In addition, AI use should aim to increase the participation of under-represented groups and reduce economic, social, gender and all other inequalities. This approach is based on a human-centred design for ‘intelligent’ or self-learning machines. It is an ethical approach to dealing with AI systems, but it is formulated very generally and neglects the legal, economic and cultural framework. Thus, some questions are not addressed, in particular concerning changes in value creation, economic structures, labour markets and employment.

*Policy approaches have a strong focus on trust and acceptance of AI, but remain generalised so far.* In June 2019, the ‘Independent High-Level Expert Group on Artificial Intelligence of the EU Commission’ (AI HLEG) presented ‘Ethics Guidelines for Trustworthy AI’ for discussion. Here too, the main focus is on fundamental aspects: AI should be ‘lawful, ethical and robust’. Prevention of harm, fairness and explainability, as well as protection of privacy, transparency, non-discrimination and accountability are defined as the main criteria. Using these criteria, the guidelines aim to maintain ‘respect’ for human autonomy. Last but not least, ‘social and environmental wellbeing’ is emphasised as an objective. After all, the guidelines highlight ‘situations which are characterised by asymmetries of power or information, such as between employers and workers’ as a particular challenge.

In November 2018, the German Federal Government adopted an ‘Artificial Intelligence Strategy’ which focuses primarily on the plan to promote ‘AI made in Germany’. This also includes a general objective for a ‘responsible development and use of AI for the common good’. Such basic approaches can also be found in declarations of companies or mergers such as ‘Partnership on AI’.

In October 2019, the ‘Data Ethics Commission’ of the German Federal Government (Datenethikkommission, DEK) published similar principles, which it links, however, with requirements for dealing with algorithmic decision-making systems. In this context, it developed a graded evaluation model that differentiates between types of damage potential according to their severity and probability (‘criticality pyramid’). Following a risk impact assessment, the DEK planned to call for transparency and disclosure obligations, information rights for data subjects, authorisation procedures as well as bans on algorithmic systems depending on the respective classification. However, the DEK did not consider the special conditions that apply to the workplace.

The EU Commission recently (as of February 2020) published a ‘White Paper on Artificial Intelligence – A European approach to excellence and trust’, which covers initial regulatory options regarding ‘risky’ areas of use of AI applications, also including the workplace. Automated application procedures or ‘situations impacting workers’ rights’ are explicitly mentioned as examples for applications classified as high-risk. Whether and which regulatory approaches should be implemented at European level will become clear after the consultation phase from June 2020.

Overall, a broad political consensus is evident at both the fundamental and abstract level. So far, there has not been enough discussion of the significance of AI for the future of labour relations and for their framework conditions. To date, there are also no concrete political proposals for workplace regulation.

### 3 A guiding framework for the introduction of AI in businesses

Use of AI systems in the workplace raises new questions regarding AI logics, data handling and work design and therefore requires binding processes for the timely co-determination and participation of employees and their representative bodies. Thus, the goal is to use humanity's collective intelligence to deploy artificial intelligence in the working environment for an efficient and productive gender-sensitive work design that promotes health and learning. To this end, advisory and support services for companies, especially SMEs, as well as for parties to co-determination and employees, should also be publicly promoted (see the German Federal Government's 2018 AI Strategy).

The potential of AI applications for humane work design should be highlighted in transparent and participation-oriented operational processes. All relevant actors for these processes should begin to collaborate before the introduction of AI systems in the business in order to jointly agree the optimisation goals for company application and anticipate significant impacts for work design in good time. This applies in particular to AI applications in which employee data is processed or which even require this (examples: personalised assistance, recruiting, human-robot collaboration et al.) as well as AI applications that can result in a loss of competence and autonomy of employees or even in job losses. In this process, usability (ISO 9241-110) by the employees should be given priority and should be gender-sensitive.

When selecting and purchasing external systems, the active cooperation of employees and their representative bodies in formulating the requirements for various AI systems must also be ensured.

*One of the primary objectives should be to use AI as assistance systems in order to reduce workloads and promote Good Work.*

A joint agreement on the objectives is of particular relevance for AI systems in order to achieve the necessary acceptance for use in the business – even in companies without a works council. This applies to both automation processes (algorithmic or automated decision-making; ADM/AuDM) and assistance systems (decision support systems; DSS).

Privacy and data protection are inherently important values which must be safeguarded. This applies to the processing of personal and personally identifiable data. But there is much more to AI systems in the workplace: AI systems have the potential to significantly change working conditions and social relations in the workplace.

One of the primary objectives should be to use AI as assistance systems in order to reduce workloads and promote Good Work. The focus must be on the occupational health and safety of employees while their personal rights are preserved at the same time. However, the broad spectrum of assistance systems alone, ranging from simple support (e.g. driving assistance) to performance-controlling or even de-qualifying systems, already suggests that the applications' modes of action should be divided into different levels of criticality. This is because AI applications also pose new risks regarding the surveillance and evaluation of employees and workforces – for example through individual profiling, company network analyses ('people/workplace analytics') or self-controlling mechanisms (from gamification to scoring). In addition, who defines the influencing parameters of assistance and support systems must be disclosed and co-determined, since, in the case of platform-based applications, only the providers have access to these parameters. Such a form of efficiency optimisation can generate additional pressure on employees, create new psychological stress and ultimately trigger new conflicts. As different levels of algorithmic decision-making or decision support are possible, the scope of action for employees and the corresponding responsibilities must be regulated.

*Using data is not just about complying with legal limits, but about developing 'Good Work by design'.*

In addition, labour law consequences for employees which could theoretically result from 'digital management' or surveillance, predictive analysis and self-optimisation mechanisms should be strictly excluded. Failing this, acceptance issues and legal infringements could become a serious obstacle to the implementation of AI systems in the workplace even if ergonomics were improved.

Therefore, using data is not just about complying with legal limits, but about developing 'Good Work by design'. In addition to determining objectives, this includes transparency for comprehensible and verifiable information about the system as well as the explainability of its mode of operation and interdependencies. This also includes the question as to the purpose for which personally identifiable data – and even anonymised data – is to and may be used. Responsibilities should be defined and strict purpose limitations and impact assessment should be implemented.

The following 'key questions' of this concept paper should not be regarded as a checklist, but as an orientation for a new understanding of business negotiation processes in order to use AI applications for Good Work.

## **Key questions for the use of AI in the workplace**

In view of the special conditions of AI use in the workplace, a comprehensive and preventive design process must be selected in which employee representative bodies and the employees involved or affected are actively involved from the outset. Various steps are recommended for the 'change impact plan', which, in addition to the definition of objectives and the formulation of requirements for the AI system providers, particularly include rules for the use of personally identifiable data and a business assessment of qualitative and quantitative consequences for work in the business with the aim of humane work design.

## Step 1 – Defining objectives and conflicts of objectives

There is a broad political and also scientific consensus that open and participatory processes are advantageous for the acceptance and thus the successful use of AI systems in the workplace. The process begins with a joint definition of the objectives, risks and limits associated with the AI application. To this end, a process agreement for negotiation and participation processes on a preventive basis should be pursued.

### The crucial questions here are:

- What are the objectives of using AI applications in the workplace?  
Are conflicts of objectives and risks foreseeable? How are these taken into account in the evaluation process?
- Are the agreed objectives for AI use documented in a comprehensible and verifiable way?
- Which employees or groups of employees are affected?
- Are the employees, groups of employees and their legal representative bodies involved in the definition of objectives and in the planning regarding the use and purpose of the AI application, as well as any employee data to be used?
- Do employees have the necessary skills to use, assess and interpret the AI system's mode of operation?
- Has it been ensured that, before the introduction of the AI application, the employee representative bodies will be advised and qualified in such a manner that they can adequately exercise their rights of participation and co-determination from the very beginning?
- Is the expertise of the business being utilised by involving the employees and can external expertise be called upon?

## Step 2 – Requirements for AI providers and developers

### *Measures must be taken to ensure that humans can continue to control and design the systems.*

The question of trustworthiness does not only refer to the functionality, robustness and security of AI systems, but also to the providers of such systems for the workplace. In addition to the overarching basic principles, special requirements regarding AI providers apply for which a certification procedure is required. Measures must be taken to ensure that humans can continue to control and design the systems. When designing and using autonomous and self-learning algorithms, human decision-making authority and human supervision have the highest priority. In the 'Hambach Declaration on Artificial Intelligence' (2019), the data protection commissioners of the federal and state governments developed proposals to define providers' obligations regarding transparency, verifiability and data protection. These include:

### Basic principles

- Compliance with legal requirements, democratic values and principles by AI providers/developers (Basic Law, Charter of Fundamental Rights, General Data Protection Regulation, Federal Data Protection Act, Anti-Discrimination Law, etc.)
- Adherence to and promotion of ergonomics and gender-sensitive occupational health and safety
- Widening the employees' scope of action
- Promotion of gender equality and diversity

### Transparency requirements for AI providers and developers ('instruction leaflet' for business use)

- High data quality and data security (also for contract data processing)
- Non-discrimination
- Usability
- Explainability (DSGVO Art. 4)
- Transparency of influencing variables/criteria and 'translation' of the algorithmic decision-making criteria and the AI's functionality
- Verifiability, traceability and correctability of decisions of the AI application
- Accountability
- Impact assessment possibilities

After defining objectives (step 1) and transparency regarding the AI application's mode of operation (step 2), criticality levels for the use of AI should be defined in order to exclude possible risks (and thus, if applicable, applications) or to agree on the corresponding measures of preventive work design according to the principle of 'Good Work by design'.

### Step 3 – Process transparency for data usage

When using AI applications, the question of whether to use personally identifiable employee data, and for what purpose, is of decisive importance in addition to the objective. Since, in addition to performance and behaviour monitoring, AI systems also offer the prospect of predictive analyses of employees, provisions must be established and recorded in an early, transparent and binding participation-based agreement as the basis for the business's use of personally identifiable employee data, as well as determining the AI objectives and functionalities. Transparent rules on data processing and business use, analysis and evaluation possibilities and their options for inspection as well as an impact assessment should be determined in advance. According to the General Data Protection Regulation (GDPR), the principle of 'Privacy by design' (Art. 25 GDPR) applies to ensure the protection of personal rights in technology design.

According to Art. 35 GDPR, a data protection impact assessment must be carried out for systems that pose a high risk to the rights and freedoms of data subjects (in this case the employees). AI systems may only be used for specified, clear and legitimised purposes (Art. 5 GDPR) and need a legal basis (Art. 6 GDPR) – collective agree-

*In the case of AI systems, extended processing purposes must also be compatible with the original purpose of data collection.*

ments, works or service agreements can constitute a legal basis in Germany under Section 26 of the Federal Data Protection Act (BDSG). Changes of purpose are clearly limited by Art. 6 (4) GDPR. It should be noted that there are stringent requirements for voluntary consent as a legal basis in the employment relationship (BAG 11.12.2014 – 8 AZR 1010/13). In the case of AI systems, extended processing purposes must also be compatible with the original purpose of data collection. The practical benefit (e.g. improving working conditions, development opportunities, easing workloads) for the employees must be made apparent and usable, which the GDPR also stipulates (mutual benefit as a basis for consent in the employment relationship).

- Is personal and personally identifiable data processed, i.e. data that makes individuals identifiable?
- Is the use of personal data necessary for the planned achievement of objectives (lawfulness, purpose limitation)?

- Are employees sufficiently informed about the use of data? Are the data subjects' rights of access (GDPR) regarding the purpose of processing, data categories, recipients, storage period, erasure concept, scope, etc. being respected?
- Has it been agreed, transparently and in accordance with collective bargaining law, in which manner (storage, duration) employee data is to be used? Also, who has access to this data and to the data-based analysis facilities/results ('participation-based collective usage agreement' as the basis for consent) at what time and in what manner?
- Has the anonymisation (aggregation) of employee data been agreed?
- Are consequences under labour law due to the use (evaluation) of employee data in the business excluded by appropriate agreements?

## Step 4 – Assessment of the consequences and humane work design in businesses

AI applications change work processes, structures and tasks, which results in great uncertainty and concern among employees regarding their personal future (job security) and interactions with 'intelligent' systems (interactions between humans and AI). This also begs the question of which qualification requirements AI triggers and how it will change the workload profiles for employees. It is important to assess these issues as preventively as possible and to continue evaluating them in the further implementation process. Here too, the issues of changes to jobs (job security) and of tasks, qualification requirements, changes in workload and data use are decisive. A criticality grid covering the different aspects can be developed for this purpose.

*How will the scope of action for employees change? And will new (physical and mental) hazards arise for employees in the business?*

- Which processes and jobs are directly or indirectly affected and in what manner?
- Which employment categories are affected and how will the employment situation in the business change?
- How will the requirement and workload profiles of jobs change?
- Are qualification measures (for and because of AI applications) necessary?
- Is there sufficient (additional) time for qualification for and familiarisation with AI applications?
- Are retraining measures necessary for employees due to AI use? What are the requirements for professional reorientation (work identity, pay classification, etc.)?
- How will the scope of action for employees change (division of labour/human-technology interface)?
- Will new (physical and mental) hazards arise for employees in the business?
- Are performance and behaviour analyses possible and/or planned to coincide with the AI use?
- Are predictive personality or development analyses of employees possible or planned?
- How is value neutrality and transparency of AI deployment ensured? How is algorithm-based discrimination and its lack of transparency prevented?
- Is the establishment of an AI complaints board planned, the composition and functioning of which will be determined by the employee representative body?



## Step 5 – Autonomy and responsibility regarding the use of AI in the workplace

AI applications can extend or limit the autonomy and scope of action of employees at their workplace. Therefore, the use of AI in the workplace imposes high standards for healthy and humane work design and also poses new questions regarding responsibility, for example when dealing with AI-based decision proposals. In particular, the use of AI must not completely displace the human decision-making prerogative in labour law issues. This applies in particular to (a) the exercise of the right to give instructions by the employer or superior, (b) the preparation and implementation of labour law measures (from warning to termination) and (c) the initiation of the employment relationship, application procedure or personnel selection. In these cases, it must always be a human being who makes the final decision.

*To what extent do AI decision proposals have a binding effect on employees?*

- According to which criteria is the interface between AI application and employees to be designed (division of labour and key responsibility/decision-making authority in dealing with AI systems, autonomy levels of human action)?
- What regulations are there regarding labelling requirements when dealing with AI applications?
- How is the work control (instruction) by AI limited (§ 106 GewO)?
- What implications does the use of AI systems have for liability issues in labour law? To what extent do AI decision proposals have a binding effect on employees?
- Which conflict resolution mechanisms and responsibilities/liabilities exist in dealing with AI systems, especially for system weaknesses?
- How is the use of AI in labour law measures (ranging from preparation of warnings to terminations) limited/regulated (KSchG)?
- How is the non-discriminatory nature of AI applications ensured (for example in job application management/personnel selection)?

## Step 6 – Tests and controls

The use of AI as learning systems also requires 'learning businesses' – i.e. test phases, interventions, an evaluation and, if necessary, readjustments. The experiences of employees and their representative bodies as AI users should be explicitly involved and formats should be chosen that include an assessment of objectives and risks. Co-determination plays a central role in this.

- What are the plans for the introduction? Will there be experimental and test phases?
- How will the experimental and test phases be evaluated and documented (change and design requirements for (a) skills development, (b) workload profiles and (c) scope of action for employees)?
- How are the opportunities for employees and their legal representative bodies to have a say in and participate in shaping the experimental and test phases defined?
- How will a transparent feedback system be organised with the involvement of the employees concerned (usability, target benefits, workloads, key responsibility, etc.)?
- What are the possibilities of intervening in and adapting the use of AI for employees and their legal representative bodies (establishment of a complaints board or similar)?

## 4 Developing participation and co-determination processes further

Co-determination processes are indispensable for AI use in the workplace, but so far, they have been limited and mostly related to specific occasions instead of being process-oriented. The introduction and use in the workplace of ‘technical devices designed to monitor the behaviour or performance of the employees’ is subject to co-determination under the Works Constitution Act (Section 87 (1) No. 6 BetrVG). Employee committees must also play a part in determining ‘the introduction and use of technical devices designed to monitor the behaviour or performance of the employees’ (Section 75 (3) No. 17 of the Federal Staff Representation Act, BPersVG). Works councils and employee committees also have the duty to ensure that the rights to privacy and informational self-determination are protected.

Even the possibility of surveillance by new technical systems is decisive here. This applies in particular to the possibilities of a systematic, data-based analysis of employees’ personal characteristics (profiling). Since the handling of personal and personally identifiable data – not least because of the predictive possibilities offered by AI – could be a sticking point for the successful introduction of AI systems, this sensitive issue must be given special attention during the implementation of AI in the workplace. The possibility of personalising non-personal data, which can only be recognised in context, is a specific challenge for AI applications with access to mass data. For this reason, co-determination must be extended to usage forms of data and techniques in order to exclude risks and to use the data and analysis potential for Good Work. This point must also be considered, not least when planning solutions to be developed externally. A high degree of transparency and binding agreements on the analytical possibilities and limits are needed. Co-determination should therefore be more process- and participation-oriented.

*Given the dynamic nature of AI applications, ‘flexible agreements’ to ensure regular process adjustments should be considered.*

Pilot and test phases to check the interfaces of the AI application with regard to the objectives and to rule out unwanted effects are also advantageous here. It is important for possibilities for intervention and correction of AI usage to also be offered in the further process, and these must

have options for co-determination. As acceptance of AI usage is therefore also increased during implementation and evaluation, co-determination should be institutionalised for the entire process chain. It is conceivable – beyond the provisions of Section 80 BetrVG and Section 68 BPersVG – that existing co-determination structures could be supplemented with both in-house and external (scientific or trade union) expertise and process support. Given the dynamic nature of AI applications, ‘flexible agreements’ to ensure regular process adjustments should be considered. This can be a basis for cooperative change management on equal footing, in that the agreed process steps are jointly reviewed and evaluated with regard to their effects and, if necessary, adapted based on learning experiences.

On the basis of the current legal framework alone, there are already options for the design of AI systems:

**In the workplace:**

- Agreement on the establishment of procedural participation and co-determination structures (to define objectives, levels of criticality, change processes, limits and revisions)
- Advisory services and gender-sensitive competence development for employees and employee representative bodies within the business
- Implementation based on participation and processes, e.g. via early, transparent and binding participation agreements (also as a basis for the employees concerned to give voluntary consent to the use of their data by the business)

**At the level of social partners/parties to collective agreements:**

- Promotion and agreement of procedural participation and co-determination structures at business level – for example through general collective agreements
- Promotion of qualification and skills development for AI applications (e.g. collective agreements for further training in areas such as safety and security, datafication, data impact assessment, basic knowledge for critical and constructive evaluation and monitoring, etc.)
- Contribution to the development of industry- or sector-specific standards (guidelines, criticality standards, etc.) for AI applications and their use
- Promotion and application of best practices for AI applications and the (sector-specific) transfer from AI and labour research

## 5 New policy requirements

The guiding framework shows that preventive and sustainable negotiation and participation processes for the use of AI in businesses are demanding and complex. Nevertheless, they are essential for safe AI usage and acceptance and the adequate protection of employees. The current legal framework offers design options to actors in the field of labour. However, these options only partly meet the requirements for the use of AI applications in the workplace. Current practice shows that there is uncertainty and that co-determination is not sufficiently process-oriented. The German Confederation of Trade Unions therefore proposes an amendment of the regulatory framework to help establish a culture of innovation based on trust and security.

# 10-point plan for a legal regulatory framework for reliable AI usage

1. **Establishment of a certification mechanism incorporated into law and development of independent bodies for testing and complaints to ensure democratically legitimised supervision and control**
2. **Promotion of business negotiation processes by strengthening co-determination rights**
  - a) in the initiation, planning and implementation of communication and participation processes of employees as well as for binding process design and the controlling of processes as part of matters subject to co-determination ('procedural co-determination right')
  - b) by extending the right of initiative and co-determination in the implementation of in-house training measures to a general right of initiative and co-determination, also during the introduction of vocational training/qualification in the business (as it is expected that the use of AI systems will lead to long-term, fundamental changes in business qualification requirements)
  - c) through a general right of co-determination for the use of personal and personally identifiable data and for the protection of the personal rights of employees
  - d) combined with facilitated access to external expertise which is not dependent on prior agreement with the employer
3. **Promotion of skills development of works councils and employee committees for the use of AI in businesses (complexity of AI systems as well as awareness-raising and qualification regarding datafication policy), e.g. through government-funded qualification and advisory services**
4. **Introduction of an independent Employee Data Protection Act to meet the special requirements for processing personal data in businesses**
5. **Prohibition of the use of legal submissions of fact and of the use of evidence for unlawfully obtained employee data and its usage**
6. **Concretisation and improvement of existing provisions of the General Equal Treatment Act (e.g. Sections 3, 12 and 15 AGG) to protect employees from algorithm-based discrimination**
7. **Ensuring the binding nature of the implementation of processes for impact assessment and evaluation of AI applications (analogous to data impact assessment according to GDPR) with regard to the change in the workload situation in the workplace**
8. **Binding nature of statutory mental health risk assessment and its adaptation to AI systems (especially anti-stress regulation, strengthening of supervision)**
9. **Expansion of labour research and critical datafication research to promote the implementation of AI projects on the basis of social partnership (development of binding standards/framework conditions)**
10. **Development and introduction of ethical guidelines in training and a Hippocratic oath for AI development negotiated on the basis of social partnership with regard to the implications of AI systems on labour and social policy, as well as support measures for business orientation along ethical guidelines**