Trustworthy AI In Human Resources

Findings of the discussion in the workshop on 18.01.2022

leiwand.ai

CHALLENGES	OPPORTUNITIES
• **Overestimation of the possibilities of AI**, since AI is often not understood or the limits / restrictions of this technology are not realised, and thus the true application possibilities and results are overestimated. • **Compliance with legal framework (data protection, liability, etc.)**. These can limit the development but must not be left to the market. Currently required proofs according to e.g. fairness are difficult to provide. • **Definition of responsibilities** for the fulfillment of various requirements. Several different perspectives must be considered: Development, vendor or user side. • **Lack of knowledge** among developers and researchers on legal and ethical guidelines to make qualified decisions on what is fair, etc. • **Lack of skills** to deal with AI based software/systems or to interpret data. Lack of training, information, or education. The requirement profile for decision makers needs to be adapted. • **Difficult perception of bias** also in society (making the subconscious aware). • **Lack of transparency** regarding algorithms, training data, data protection, etc., decisions must be comprehensible. • **Algorithmic discrimination** by AI systems if not all criteria, biases, etc. are properly considered. • **Lack of data quality**, because biased data / poor quality data are used as a starting point for an AI development. This leads to unknown secondary biases, among others. • **Access to unbiased data** is almost impossible due to lack of data, but also bias in society. • **Lack of access to data** for research and development. • **Lack of knowledge about needs** and intended uses, so AI is developed in ignorance of market needs. • **Lack of co-determination of all stakeholders** in the development and use of AI (incl. veto rights). • **Human decision-making/control** of big data is very difficult. • **Lack of measurable / addressable criteria** for quality control. • **The impact of human intervention in AI decisions has not yet been clearly researched**. For example, it is still unknown when meaningful human control exists (automation bias), and when potential biases arise from such intervention. Lack of documentation / traceability when humans intervene in AI decisions. • **Consideration of a company's / organization's values** in AI-supported decision-making processes is difficult to implement. • **Lack of trust in AI-supported decisions**, especially skepticism or lack of acceptance by employers and employees. With automation, there needs to be a lot of trust. • **Risk of discrimination**, that AI systems support people in already privileged sectors, but discriminate more in less privileged sectors. • **Lack of openness to the future**, so that the possibility of the completely different is not given.	• **Existing discrimination patterns become visible,** and this creates a great opportunity to question previous decisions and to take action against discrimination in society, to reduce or prevent it. • **Disclose decisions and document them** in a transparent and comprehensible way. Human factor / informal decisions and structures have reduced influence. • **Control of decisions** is made possible. • **Ethical discussion** is stimulated and enables new opportunities. • **Initiation of discussions** on the values and conditions of the use of algorithms. • **Strengthening diversity** as well as increased fairness through data-based analysis. Subjective judgments are minimized, discrimination is prevented or reduced by focusing on qualifications. • **Increased fit** between employees or applicants and thevacancy to be filled. • **Opening of the job market and equal opportunities** are promoted, e.g. equal access to information. People learn about positions or are considered for them who would otherwise never have been interested in or applied for the position. • **Process optimization**, since processes can become faster and cheaper or easier to use. This results in economic advantages, such as the reduction of time and financial resources. • **Labor relief and a good supplement to human-managed processes** "AI as an additional employee". However, the human being is not completely replaced. • **Increased quality of work** by, for example, recognizing and preventing demotivations at an early stage and thus achieving more joy at work, less sick leave etc. • **Making hidden champions transparent** and promoting the use of previously unused talents. • **Individualization**, e.g. in career paths, will be increased (career on demand). • **Counteract shortage of skilled workers** to some extent. Fairness is in demand and new target groups can be tapped by breaking down previous barriers. • **Employer branding** can benefit from the use of AI if it is communicated transparently internally and externally, and the limitations are openly discussed. • **Organizational resilience** and engagement. Operating fairly also means operating resiliently. • **Large amounts of data can be analyzed** through standardized processing and better handling of highly complex issues is made possible. This enables new insights. • **Stakeholders and advocacy shape top to bottom** and can drive the issue forward. • **Universal regulations and certifications** enable Trustworthy AI and its opportunities.