

Quality Criteria for AI Development and Implication

Human rights values, such as freedom of action, equal rights, solidarity, diversity, participation and pluralism of values, are to be strengthened, not restricted, by algorithmic systems. To ensure this, we need quality criteria for the development of AI/Algorithms.

The following quality criteria are aimed at all persons who have a significant influence on the emergence, development and programming, use and effects of algorithmic systems. The focus is on those algorithms that have a direct or indirect influence on people's lives. They are based on the recognition that the course for beneficial AI is already being set during development and cannot only be achieved by subsequent regulation. The following quality criteria are not a substitute for regulation. Their early consideration and implementation enables smart regulation in the first place.

Quality Criteria for beneficial AI:

Build-up competence

The functions and effects of algorithms must be understood. Those who develop and deploy algorithmic systems must have a clear understanding of the functioning and potential impact of the technology. The transfer of individual and institutional knowledge, but also the interdisciplinary exchange within organisations, are essential.

Human Responsibility

A natural person must be responsible for the effects of an algorithmic system. For this it is necessary that the person in question is aware of his or her responsibility and the associated tasks. The responsibility assigned to him must be documented and the person in question must be recognizable from the outside. At all times, they must be able to reverse processes and make new decisions. Neither can a machine assume this responsibility, nor may liability be passed on to the end user or affected person. If there is a shared responsibility, for example by several persons or organisations, the usual legal regulations on the assignment of responsibility apply.

The goals and expected effects must be made comprehensible

The goals and objectives must be clearly defined in advance and information on the use of the algorithmic system must be documented. This includes, for example, the underlying data and calculation models. A documented impact assessment must be carried out before and during the use of the algorithmic system. Risks for discrimination and other consequences affecting the individual and the common good must be considered. Value considerations and balancing in the setting of objectives and the use of algorithms must be recorded.

Ensure security

The use of algorithmic systems must be secure. Before algorithms are used, they must be tested in a protected environment. The technical security against external attacks and manipulations as well as the security of the users must be guaranteed. Unsecure or untested systems should not be rolled out.

Integrated transparency

The use of an algorithmic system must be labelled. This includes making algorithms and self-learning systems with direct or indirect effects and their functionality intuitively understandable for humans. The underlying data must be classified and described and the possible effects presented in easily understandable language. Persons affected by the effect of an algorithm can demand qualified and detailed information on these parameters. When a machine imitates a human being in speech and in the way it interacts, this is particularly true. Anthropomorphization (Humanization of technology) should be avoided.

Design and implementation of controllability

The use of algorithmic systems must be controllable. This presupposes that people are able to decide on the algorithm and its use and to evaluate it according to ethical standards. A constant control of the algorithmic system by humans is necessary to ensure controllability. This applies in particular to self-learning systems. If it is not possible to make an application controllable for humans, an application should not be used.

Permanent impact monitoring

The effects of an algorithmic system on humans must be checked regularly. This includes active monitoring of whether an algorithm violates basic social values such as value pluralism, solidarity, diversity and participation. External independent experts should also be enabled to verify this. If a negative effect is detected, the cause of the error must be determined and the algorithm adapted accordingly. If the error cannot be corrected, the use of the algorithmic system must be terminated.

Establishing Correctability

Decisions of an algorithm must never be irreversible. If the effects of an algorithm result in the violation of fundamental social values or the rights of an individual, the parties concerned must be given a simple opportunity to file a complaint. The complaint is addressed to the person responsible. This person undertakes to respond to the complaint in a qualified manner and to provide information. Complaints and measures taken must be documented.

Segregation effects must be counteracted

Human contact should not become a luxury good, only for few. Therefore algorithmic/ automated handling should not be implemented as the only way of contact.

Support (medical care, customer service, etc) or evaluation by a person rather than by an algorithm, must not be determined solely by financial means.

Access to education and employment must remain fair and must not be determined solely by algorithmic decision-making.

Ramak Molavi - iRights LAB Berlin

ramak@thelawtechnologist.com

www.the-law-technologist.com