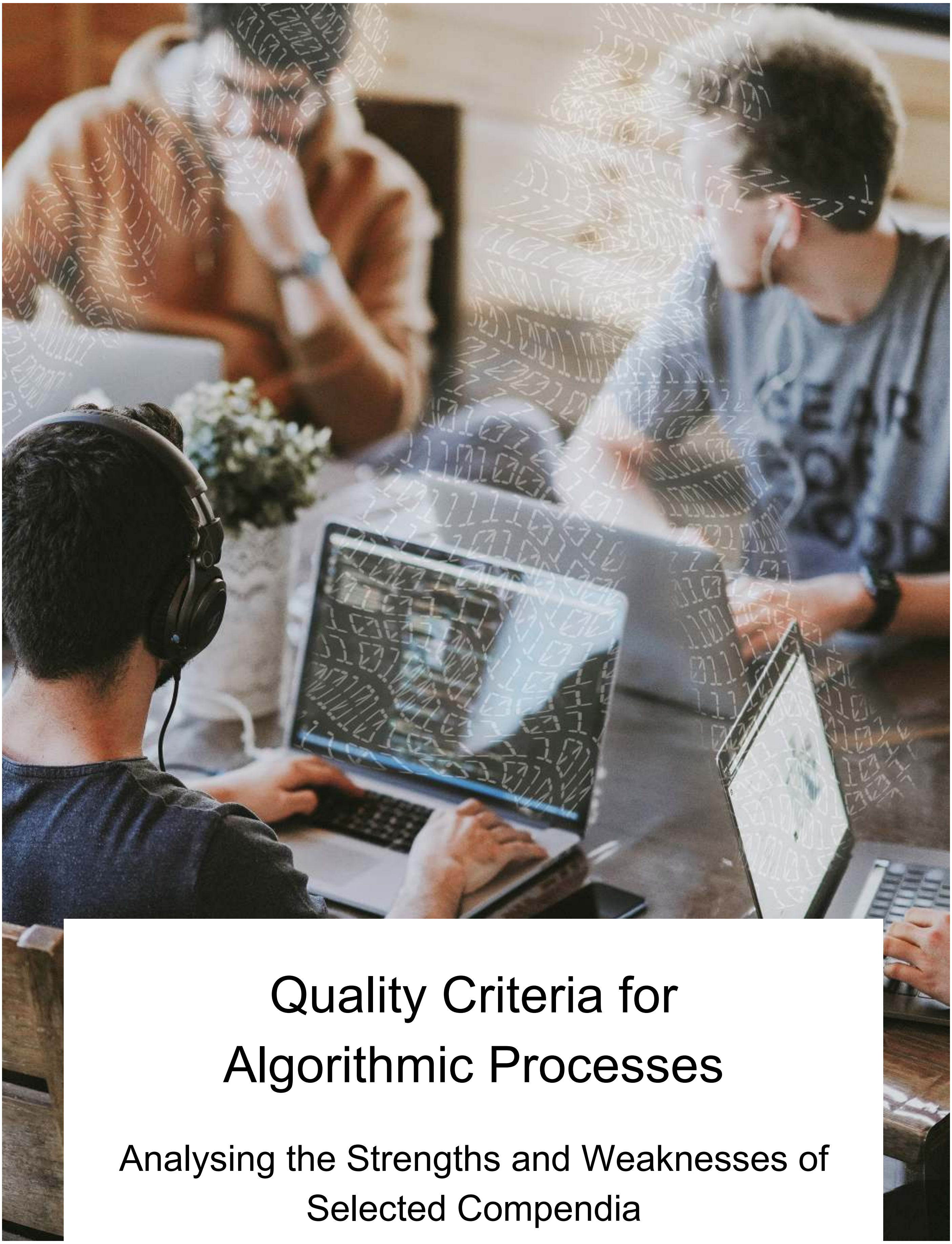
Discussion paper Ethics of Algorithms #8



Quality Criteria for Algorithmic Processes

Analysing the Strengths and Weaknesses of Selected Compendia

BertelsmannStiftung

Quality Criteria for Algorithmic Processes

Analyzing the Strengths and Weaknesses of Selected Compendia

-Working Paper-

Noëlle Rohde

Legal Notice
© July 2018 Bertelsmann Stiftung

Bertelsmann Stiftung
Carl-Bertelsmann-Straße 256
33311 Gütersloh, Germany
www.bertelsmann-stiftung.org

Responsible for content
Ralph Müller-Eiselt

Author
Noëlle Rohde

Translation
Barbara Serfozo

Cover: Annie Spratt / unsplash.com (Public Domain)

DOI 10.11586/2018028 https://doi.org/10.11586/2018028

Contents

1 Preface

What should algorithms be allowed to do and what not? What standards of quality should they be held to? For what purposes may they be used? These are relevant questions for all of us as individuals. As a society, we need to achieve a consensus in our response to these questions. The people behind algorithmic processes, that is, those who design these processes, are the most important factor in determining an ethics of algorithms. Everyone from clients to programmers to information and data scientists must be conscious of their accountability in the creation and implementation of technology – and act accordingly.

And yet, because of the sheer diversity of actors involved in the field, it is difficult to render a specific profession accountable and make clear demands of its agents specifically. Indeed, it makes more sense to start with the algorithmic process itself – which connects all actors involved.

Several high-level international working groups, esteemed institutes and associations have already taken on the task of formulating quality criteria for algorithms. Quite clearly, these organizations have taken on a challenging task and have each invested considerable effort in producing promising results. However, these individual criteria proposals often stand independent of each other as silos – being the result of design processes that are somewhat oblivious to the lessons that could be learned from previous endeavors. This kind of silo approach fails to take advantage of potential synergies, and the resulting host of different documents makes it difficult to develop a professional code of ethics underpinned by centralized and universally accepted standards. In addition, the discussion regarding a code of ethics for algorithmic processes and their design is dominated by the international English-speaking community.

The analysis presented here compares three individual quality criteria proposals: the *Principles for Accountable Algorithms and a Social Impact Statement for Algorithms* (FAT/ML Conference), the *Asilomar AI Principles* (Future of Life Institute) and the *Principles for Algorithmic Transparency and Accountability* (ACM U.S. Public Policy Council).

This publication is designed to identify the strengths and weaknesses found in each and to highlight those criteria that are transferable to the German-speaking world. In what follows, we outline the contextual factors shaping the development of each set of recommendations and evaluate their individual content and recommendations for implementation. The analysis highlights those criteria that are particularly relevant for a German compendium of quality criteria as well as those that are less helpful. It also identifies gaps in the existing compendia that require prudent attention.

The diversity of meaningful strategies is reflected in the three proposals subject to analysis here – they are the product of research and practitioner-driven conferences (FAT/ML) or more direct policy efforts (ACM U.S. Public Policy Council), they focus on specific aspects of algorithms such as accountability (FAT/ML), or they aim to address the range of issues important to handling artificial intelligence. Each proposal aims to serve as a set of guiding principles (FAT/ML, ACM U.S. Public Policy Council) in meeting accountability design needs in algorithmic processes or addresses attitudes and assumptions towards the subject matter (Future of Life Institute).

We offer this analysis as a working paper for others to draw upon and develop in a rapidly transforming field. We are therefore happy to receive suggestions regarding additional information, improvements and, of course, constructive criticism. In order to foster ongoing discussion, we are registering this working paper under a creative commons license (CC BY-SA 3.0 DE).

The analysis presented here has been developed within the context of the Ethics of Algorithm project at the Bertelsmann Stiftung, which is tasked with examining the impact of algorithmic decision-making systems on society. The project has delivered several publications as part of its discussion paper series, including a collection of international case studies (Lischka and Klingel 2017), a study of how algorithmic decision-making (ADM) can affect participation (Vieth and Wagner 2017), an analysis of the impact of algorithmic processes on social discourse (Lischka and Stöcker 2017), a paper on error sources and accountability in ADM (Zweig 2018) and an expert opinion piece on what impact the EU’s GDPR legislation will have on ADM (Dreyer and Schulz 2018). Most recently, the project has published the findings of a survey conducted in Germany to establish what the general public understands and thinks when it comes to algorithms (Fischer and Petersen 2018) and a panorama of strategies to ensure algorithmic processes serve the public (Krüger and Lischka und Klingel 2018).

The paper featured here scrutinizes existing recommendations for quality criteria in developing algorithmic systems. Later this summer, we will be publishing another study in which we examine how to meaningfully implement such criteria with impact. This forthcoming publication will explore codes of ethics that are already established in other professions in order to identify which aspects thereof can be transferred to the field of algorithmic system development. In addition, the Bertelsmann Stiftung is working together with the Think Tank iRights.Lab as part of a broadscale stakeholder process to develop a recommended set of quality criteria that is scheduled to be published in the fall of 2018.

R. Müller-Eiselt

Ralph Müller-Eiselt
Senior Expert Taskforce Digitalization
Bertelsmann Stiftung

2 Introduction

As automated decision-making processes steadily gain influence in socially sensitive contexts, regulating algorithmic systems is an increasingly important necessity. One promising way of achieving this is through an appeal to the moral obligation of the responsible parties and the creation of a commonly held professional ethic, focusing not on general moral questions, but rather on those that arise directly from and are characteristic of the respective professional activities.

Alongside a professional identity and an implicit ethos of vocation, an additional form of professional ethic is the orientation toward an explicit body of regulations that sets out the responsibilities and duties of a particular professional group. Successful examples of this model include the Hippocratic Oath (Declaration of Geneva) for medical practitioners and the Press Code for journalists.

Defining the profession of “those involved in designing socially relevant algorithmic processes” is a difficult task, because of the heterogeneity of the actors involved. Situating responsibility solely with programmers would inevitably amount to a simplification: A multiplicity of other groups such as customers and clients, political decision-makers, and institutional representatives play significant roles in the planning, commissioning and use of algorithmic systems (see Zweig 2017) and thus influence the degree to which programmers are free to act.

One means of addressing the large variety of responsible parties is the adoption of a process-based concept of the profession, according to which everyone involved in the design of an algorithmic process, is regarded as sharing responsibility for it. Thus, developing a compilation of quality criteria that reference the process appears considerably more practical in addressing this group of people than would be the use of a code of professional ethics that seeks to guide the practitioners’ behavior directly.

In recent years, numerous organizations and working groups around the world have addressed the problem of quality assurance for algorithmic processes. Because this discourse is significantly less developed in Germany, an analysis of a selection of international proposals is useful in order to derive insights for the creation of a German-language blueprint. For the current paper, we have selected three of the most recent documents: the ACM U.S. Public Policy Council’s Statement on Algorithmic Transparency and Accountability, the Future of Life Institute’s Asilomar AI Principles (in this context, AI refers to artificial intelligence), and the FAT/ML organization’s Principles for Accountable Algorithms and Social Impact Statement for Algorithms.

Below, we describe and analyze these three quality-criteria proposals on the basis of the three following dimensions:

- Who are the authors? Under what circumstances did the proposal emerge, and what are its objectives? (Description)
- What is its content? What are the specific demands? Who is the target audience? (Description)
- What are the proposal’s strengths? Where is there room for improvement? (Analysis)

Finally, using these findings as a basis, we will identify transferable strengths and weaknesses to be corrected when developing a compendium of quality criteria.

3 ACM U.S. Public Policy Council (USACM)

The ACM U.S. Public Policy Council is a part of the Association for Computing Machinery (ACM), and is an independent body that addresses issues of computer and information technology within the U.S. public-policy context. The Council provides the U.S. Congress, administration and courts with expert assessments and analyses of current developments within its field of expertise. The U.S.-based unit frequently collaborates with other ACM policy councils around the world in developing research-informed positions, publishing and disseminating the results in the form of public-policy statements and reports. The committee for computer issues includes representatives from academia, company executives, and other experts from industry, government and the non-profit sector (ACM 2018).

On 12 January 2017, the ACM U.S. Public Policy Council published its Statement on Algorithmic Transparency and Accountability, which was subsequently also signed by the ACM Europe Policy Committee on 25 May of the same year (ACM 2017). This document contains seven criteria for algorithmic decision-making (ADM) processes aimed at increasing their benefits and minimizing any potential harm resulting from them.

As noted by the authors in the introduction, the background and motivation for the paper is the observation that algorithms are today ubiquitous and pose the risk of producing undesired social biases. For this reason, the signatories request that ADM system developers and policymakers alike observe the seven standards drafted by the ACM U.S. Public Policy Council in all contexts relating to algorithmic systems.

3.1 Description

The standards detailed by the ACM include awareness, access and redress, accountability, explainability, data provenance, auditability, and validation and testing.

- The concept of awareness alludes to the idea that owners, designers, producers, users and other stakeholders should clearly recognize that undetected biases in socially sensitive ADM processes could have negative consequences for society and individuals.
- The demands for access and redress refer to the idea that anyone adversely affected by an algorithmically informed decision should be provided with appropriate appeal and review options.
- The third demand is that institutions should be held responsible for the decisions made or informed by their algorithms.
- The standard of explainability for algorithmic mechanisms means that the responsible institutions should be encouraged to publicly describe their systems' internal processes and decision-making functions.
- Similarly, the origin of the data used to train the system should be made public, and the biases that may have resulted from the dataset should be reviewed. The maximum potential for correction can be realized only through critical public scrutiny of the data, the paper argues. However, for data-security reasons, giving only a certain qualified group of people a closer look at the data is sometimes justified.
- According to the ACM Council, the sixth quality standard, that of auditability, can be achieved only if the models, algorithms, data and decisions are recorded in such a way that they can be understood at a later date.
- Finally, the document proposes that institutions should employ rigorous test and validation procedures for their models in order to identify any possible discriminatory effects at an early date. In addition, the results of these tests should be made public.

The target audiences for the quality criteria are defined for each individual component, and range from regulators and institutions, designers, system owners and users to policymakers.

3.2 Analysis

3.2.1 Strengths

As a compendium of quality criteria for socially relevant ADM processes, the ACM U.S. Public Policy Council's Principles for Algorithmic Transparency and Accountability are in many respects well thought through and scrupulously composed. One great advantage of the ACM principles is that they are based on a complex understanding of the development of algorithmic systems. For example, the quality standards are not limited solely to the programming of the software itself; they additionally incorporate preliminary steps such as the awareness of potential problems, the collection of data and the training of the algorithm, as well as later steps such as the contestation of unfair outcomes and the ability to reconstruct and understand errors.

This multidimensional understanding of algorithmic processes is also reflected in the above-noted diversity of target audiences. While the public mind often regards software developers as holding sole responsibility for such processes, this document also assigns responsibility to policymakers, specific clients and customers, institutions and users. In other words, it addresses those who have any influence over the design of an ADM process. A process-oriented professional conception of this kind is particularly useful because it reflects the complexity of practical realities instead of simplifying, and thus running the danger that numerous relevant actors will be ignored.

Similarly, the quality standards are not limited to a focus on software systems themselves; they also encompass the data being processed, thus including an often-neglected source of error.

In addition, on a number of issues, the ACM principles show a high degree of sensitivity to potential conflicts between the quality criteria and other possible demands on algorithmic systems. For example, the demand for data disclosure is tempered so as to take account not only of algorithmic transparency and responsibility, but also of the issues of data security, business secrecy and possible system misuse.

3.2.2 Weaknesses

Nevertheless, the ACM Public Policy Council's principles also display a number of weaknesses that should be corrected in future compendia of quality criteria. First, at a purely formal level, the proposed quality standards do not follow a logical temporal order. While awareness (Principle 1) certainly refers to one of the first stages of ADM system development, the call for access and redress (Principle 2), for example, seems to be relevant only after a system has been developed and deployed. Thus, it is not immediately clear why this criterion was placed above that of institutional accountability, for instance. Similarly, software testing and validation is listed as the last point, although this should logically also be located before Principle 2. Even if this inconsistency has no further influence on the quality of the criteria, it can interfere with efforts to understand them. Using a stringently logical sequence would also help define participant groups on the basis of the temporal order of their participation in a specific ADM process.

Moreover, the principles largely neglect a particularly important aspect of algorithmic decision-making – that of ethically appropriate use. The only criteria that indirectly raise this issue, a defining feature in any ethic of algorithms, are awareness (Principle 1), which cites "the potential harm that biases can cause to individuals and society," and the demand for software testing (Principle 7) in order to determine whether a model "generates discriminatory harm" (ACM U.S. Public Policy Council 2017: 2). However, the ethical use of algorithmic systems is of such great societal importance that this issue should be represented in its own quality criterion. We will discuss below how such a criterion could be developed.

An additional weakness of the ACM principles is that their demands are all quite abstract. This circumstance is presumably due to the fact that the field of algorithmic decision-making is extremely large and diverse. It is challenging to formulate concrete criteria without unintentionally excluding a portion of the field in which they are meant to be applied. Nevertheless, an abstract catalog makes it difficult to derive practicable recommendations for action.

Grammatically, it is striking that all of the ACM's principles are formulated in the subjunctive mood (for example, "Regulators should encourage the adoption of mechanisms that..."). This simultaneously highlights one of the greatest problems with this catalog of demands: its non-binding nature. Particularly where economic or political interests run counter to the quality criteria, for example in the case of a time-intensive software validation and testing process before market introduction, the ACM principles do not clearly state why the actors involved should adhere to the proposed standards. Because the criteria are only proposals, there is neither a positive duty to follow them, nor sanctions for non-compliance. Although the establishment of such authority certainly exceeds the ACM Public Policy Council's powers, it is elements such as this that serve to distinguish an abstract compilation of standards from successful and established professional codes of ethics such as the Press Code. In particular, the document fails to address the issue of policymakers' responsibility for supporting the quality criteria's authority.

In summary, the ACM U.S. Public Policy Council's Principles of Algorithmic Transparency and Accountability represent a modern, sophisticated approach to formulating standards for ADM processes. Based on a nuanced understanding of the complexities of algorithmic decision-making, the document provides helpful orientation for the actors responsible in this area. However, it sometimes loses sight of economic and political realities, and thus does not exhibit a binding nature.

4 Future of Life Institute's Asilomar AI Principles

The Future of Life Institute is a volunteer-run research institution with the self-defined goal of minimizing existential risks. It identifies existential risks as those that could wipe out a large portion of humanity – a category that includes artificial intelligence (AI), according to the group's members (Future of Life Institute 2018a). The Institute is made up of actors from academia, public life and industry (Future of Life Institute 2018b).

The Future of Life Institute's 23 so-called Asilomar AI Principles emerged from the 2017 Beneficial AI Conference, and are named after the event's location. During the conference, academic and industry representatives worked with thought leaders from the business, legal, ethics and philosophy fields to draw up principles for the socially responsible design of ADM processes (Future of Life Institute 2017b).

The development process was initiated by the conference organizers, who compiled a list of the most commonly encountered positions regarding the management of artificial intelligence, using existing approaches to quality criteria for algorithmic decision-making as a basis. In a first revision stage, this list was provided in advance to all conference participants, with feedback and suggestions for improvement then incorporated into the draft. In the second stage, the principles collected in this way were again discussed and revised in individual groups at the conference itself. Finally, all participants were asked to approve each individual version of each individual principle. Only those principles obtaining a level of support of 90 percent or more were included in the final list (Future of Life Institute 2017c)

4.1 Description

According to the document's preamble, the following supposition serves as motivation and background for the principles: "Artificial intelligence... will offer amazing opportunities to help and empower people in the decades and centuries ahead" (Future of Life Institute 2017a). It can thus be said that the Asilomar AI Principles are based on an optimistic vision of the societal use of artificial intelligence (AI).

The Asilomar AI Principles are divided into three general areas: Research Issues, Ethics and Values, and Longer-Term Issues. Overall, the document poses 23 demands (paraphrased here from the original).

The area of Research Issues contains five demands:

1. The goal of AI research should be the creation of beneficial intelligence, not undirected intelligence (*Research Goal*).
2. Investments in artificial intelligence should be accompanied by research funding that ensures the beneficial use of AI. This should include research into complex issues from the fields of computer science, economics, law, ethics and social studies (*Research Funding*).
3. There should be a constructive and healthy exchange between AI researchers and policymakers (*Science-Policy Link*).
4. A culture of cooperation, trust and transparency between AI researchers and developers should be fostered (*Research Culture*).
5. The teams involved in the development of AI should actively cooperate in order to prevent the circumvention of safety standards (*Race Avoidance*).

The area of Ethics and Values contains 13 demands:

6. AI systems should be safe during their entire operational life. Moreover, they should be verifiably so where applicable and feasible (*Safety*).
7. If an AI system causes harm, it should be possible to determine why (*Failure Transparency*).
8. Any involvement by an autonomous system in judicial decision-making should entail a satisfactory explanation that can be audited by a competent human authority (*Judicial Transparency*).
9. The designers and builders of advanced AI systems are stakeholders in the moral implications of the use, misuse and actions of these systems. The stakeholders have both the responsibility and the opportunity to shape these implications (*Responsibility*).
10. Highly autonomous AI systems should be designed so that their objectives and behaviors can be aligned with human values throughout their entire period of use (*Value Alignment*).
11. AI systems should be designed and utilized in such a way that they are compatible with the ideals of human dignity, rights, freedom and cultural diversity (*Human Values*).
12. People should have the right to access, manage and control the data they generate, given AI systems' power to analyze and utilize that data (*Personal Privacy*).
13. The application of AI to personal data must not unreasonably curtail people's real or perceived freedom (*Liberty and Privacy*).
14. AI technologies should benefit and empower as many people as possible (*Shared Benefit*).
15. The economic prosperity generated through the use of AI should be shared widely, in order to benefit all of humanity (*Shared Prosperity*).
16. Humans should be able to choose how and whether to delegate decisions to AI systems in order to achieve objectives chosen by humans (*Human Control*).

17. The power conferred by controlling highly advanced AI systems should be used in a manner that respects and improves the social and civic processes contributing to a healthy society, rather than undermining them (*Non-subversion*).

18. An arms race in lethal autonomous weapons should be avoided (*AI Arms Race*).

The area of Longer-Term Issues contains five final demands:

19. Because there is no consensus, absolute assumptions about the upper limits of the future capabilities of AI should be avoided (*Capability Caution*).

20. Advanced AI could mean a profound change in the history of life on Earth and should therefore be planned for and managed with appropriate care and sufficient dedication of resources (*Importance*).

21. Risks presented by AI, particularly of a catastrophic or existential nature, must be subject to planning and mitigation efforts commensurate with the expected level of risk (*Risks*).

22. AI systems that recursively self-improve or self-replicate in such a way that their quality or quantity would rapidly increase must be subject to strict safety and control measures (*Recursive Self-Improvement*).

23. Superintelligence should be developed only in the service of widely recognized and shared ethical ideals, as well as for the benefit of all humanity, rather than for an individual state or individual institution (*Common Good*).

4.2 Analysis

4.2.1 Strengths

The 23 Asilomar AI Principles together constitute a multifaceted catalog of proposals for the ethical design of ADM processes. For example, their complexity is clear in the fact that the criteria relate not only to the substance of algorithmic systems, but also take account of their genesis and far-reaching ramifications.

As a direct consequence, one of the greatest strengths of the Asilomar AI Principles is their inclusion of the area of research in their considerations. In addition to setting an ambitious research goal (the creation of *beneficial intelligence*), the document takes research funding specifically into account. As already noted, this should facilitate research into the socially beneficial use of artificial intelligence. The principles explicitly cite not only computer science in this context, but also economics, law, ethics and the social sciences. In this way, the authors of the Asilomar Principles succeed in involving a variety of actors who are not ordinarily regarded as being responsible for the development of ADM processes.

An additional virtue of the Asilomar Principles lies in the fact that they discuss meta-issues relating to the role of artificial intelligence as well as making practical demands. These broader issues are well described, particularly in the *Ethics and Values* section. In this regard, it is worth emphasizing that the values identified here go beyond aspects already widely discussed in public such as transparency, privacy, security and responsibility. In so doing, the principles incorporate deeper fundamental tenets such as the use of artificial intelligence for the purposes of *shared prosperity* and *shared benefit.* The catalog's inclusion of two quality criteria specifying algorithmic systems' congruence with human values (principles 10 and 11) is of particular importance in this context. Here, Principle 11 can be seen as an intensification of the content of Principle 10. The concrete way in which some of these value propositions are anchored is certainly one of the Asilomar Principles' unique features. For example, the demand that algorithmic systems be made compatible with the ideals of human dignity, rights, freedoms and cultural diversity is an important reminder that artificial intelligence holds great social relevance and cannot be allowed to contravene societal values.

Particularly in the context of societal ideals, the Asilomar Principles resemble a legislative text that sets high goals, but also offers opportunity for interpretation. One advantage of this approach is that the criteria can thus be applied to a wide variety of artificial-intelligence uses, instead of being relevant solely within specific fields. Possible disadvantages of this legislation-like style of demand will be discussed below.

The Asilomar Principles suggest that the authors have a deep understanding of the power of algorithmic processes. This is clear in the way the document takes account of the ramifications for fundamental social values, as well as in the inclusion of a special section of demands and assumptions addressing long-term problems. Particularly important is the reminder that advanced artificial intelligence could mean a profound change in life on Earth, and that its use could simultaneously produce catastrophic and existential risks. This explicit statement, highlighting the far-reaching influence of algorithmic systems, serves to recursively emphasize the imperative nature of the entire catalog of demands.

Finally, the Future of Life Institute’s Principles are quite nuanced. With 23 separate hypotheses and demands, the collection is by some distance the most comprehensive of the proposals analyzed here. For example, instead of the often-heard call for transparency in algorithmic systems, it distinguishes between *Failure Transparency* and *Judicial Transparency* (see above for details). An approach of this kind enables the parameters of selected ideals to be more precisely defined and is partially responsible for the Asilomar Principles’ resemblance to a legislative text.

4.2.2 Weaknesses

One of the most serious drawbacks of the Asilomar AI Principles is their failure to identify target audiences for the quality criteria with sufficient specificity. Some of the demands make it clear who they are oriented toward; in the case of the Science-Policy Link (Principle 3), for example, the target audience is largely researchers and policymakers. However, this is only rarely the case. The *Ethics and Values* and *Longer-Term Issues* demands in particular lack any reference to possible addressees or to the parties ultimately responsible for observing the quality criteria. Even designers and builders are mentioned in only one place (Principle 9), identified in this context as stakeholders in the moral implications of artificial intelligence.

Given the great diversity in the use of algorithmic decision-making systems, any such limitation to computer scientists as the sole responsible parties is not tenable. Rather, a multiplicity of actors directly or indirectly involved in the design of socially relevant algorithmic processes must be taken into consideration. For example, researchers and policymakers should be addressed not only in the context of research issues, but also when discussing ethical criteria for artificial intelligence. Moreover, other important stakeholders such as public institutions, customers, clients and users go virtually unmentioned as addressees. This lack of sensitivity for the diversity of responsible parties results in an insufficiently direct appeal; this in turn poses the risk that many relevant actors will fail to perceive the Asilomar AI Principles as being relevant to them.

In a similar way, the principles’ authors also wholly overlook the public discourse as a possible tool for the ethical design of artificial intelligence.

As already noted, the Asilomar Principles are quite open in their formulations. Although this facilitates great flexibility with regard to their application – something that is certainly helpful given the complex nature of algorithmic processes – the criteria often risk being both too general and too vague. For example, the document calls for an exchange between AI researchers and policymakers that is “*constructive and healthy*.” Without the specification of any parameters for this statement, it is difficult for any such criterion to serve as an ethical guideline, particularly because definitions of “constructive” and “healthy” depend greatly on the perspectives of those involved.

Similarly, the Asilomar Principles' authors often fail to provide details regarding the practical implementation of their demands. For example, one item indicates that AI development teams should actively cooperate in order to avoid a competition-driven erosion of safety standards. However, it is unclear how any such admonition could become a guiding principle for action if economic interests were to stand in opposition to the ethical demands. At this point, the Asilomar Principles could benefit from a reference to specific political measures.

In short, the Future of Life Institute's Asilomar AI Principles are a highly sophisticated catalog of quality criteria that show great sensitivity to the most important fundamental and practical issues in the area of ADM. The criteria go well beyond demands for algorithmic systems themselves, and even indirectly include actors (such as social scientists) who are often overlooked in discussions of the social design of algorithmic decision-making processes.

At first look, some of the demands might seem rather idealistic and even somewhat unrealistic, particularly as the document proposes no tools or practices to facilitate their implementation. However, one can also argue that a collection of quality criteria of this kind should be understood in the same manner as a legislative text. In this case, this means setting high societal ideals, while still offering sufficient interpretive flexibility to do justice to individual cases in the field of algorithmic decision-making.

5 FAT/ML Principles for Accountable Algorithms and Social Impact Statement for Algorithms

FAT/ML stands for Fairness, Accountability and Transparency in Machine Learning and is also the name of an annual conference that offers scientists and practitioners in relevant areas the opportunity to address and discuss societal challenges in the field of self-learning machines. The conference, which took place for the fourth time in 2017 (FAT/ML 2018), is organized by a permanent committee of scientists from numerous universities as well as from Google and Microsoft Research.

The paper "Principles for Accountable Algorithms and Social Impact Statement for Algorithms" is listed as a background document on the FAT/ML website and is the result of the Dagstuhl Seminar "Data, Responsibly" from 17 to 22 June 2016. The 39 participants in this seminar (including ten women) consisted of scientists from a range of international universities and institutes – in particular from the North and South American as well as European regions – alongside representatives from Google and Microsoft. The participants represented numerous research fields in computer science, including data management, data mining, security and data protection and computer networks. Social scientists, data journalists and think tank representatives also took part in the process. The paper on Algorithmic Accountability was signed by thirteen of the participants (FAT/ML 2016).

The background to the Dagstuhl Seminar was a recognition of the wide-ranging influence on society exerted by big data and related technologies, as well as a demand for this influence to be applied responsibly and in accordance with general ethical and moral norms. According to the seminar organizers, in order to come closer to this ideal, there is an "urgent need to define a broad and coordinated computer science research agenda in this area" (ibid.). The declared aim of the Dagstuhl Seminar was the development of a first draft of this agenda – the result being the Principles for Accountable Algorithms and a Social Impact Statement for Algorithms.

5.1 Description

As indicated by the title, the FAT/ML document is divided into two sections: the Principles for Accountable Algorithms and a Social Impact Statement for Algorithms.

The Principles are preceded by a preamble and a premise. The former includes the observation that the prevalence of automated decision-making processes is increasing steadily, and that algorithmically "informed decisions are now capable of exerting considerable social influence.” (ibid.). The Principles for Accountable Algorithms aim to help developers and product designers in the design and implementation of algorithmic systems that are publicly accountable. In this context, the concept of “accountability” refers to the obligation to document, declare or justify algorithmic decision-making processes, and to mitigate any related negative social consequences.

The premise exists mainly the proposition that algorithms and the data that they utilize are human creations. This means that for every automated decision, there is ultimately an accountable human being, and thus the excuse of “it was the algorithm” cannot be validly applied.

On the basis of this premise, five Principles for Accountable Algorithms are presented:

- *Responsibility:* Externally visible opportunities have to be created to rectify effects that are adverse either for the individual or for society as a whole. In addition, an internal role for the respective responsible person to quickly correct issues of this kind must be established.

- *Explainability:* It must be ensured that algorithmic decisions as well as the underlying data basis can be explained to end users and stakeholders in non-technical language.

- *Accuracy:* Sources of error and uncertainties on the side of the algorithm and the underlying data basis must be identified, logged and articulated in such a way as to facilitate both the understanding of expected and worst-case scenarios and the development of appropriate countermeasures.

- *Auditability:* Interested third parties must be given the opportunity to scrutinize, understand and verify the behavior of the algorithm. This should be carried out through the release of information that permits the monitoring, examination or critique of the system. Such information includes, among others, detailed protocols, technically adequate APIs (application programming interface) and permissive terms of use.

- *Fairness*: It must be ensured that algorithmic decisions have no discriminatory or unjust effects in the comparison of different demographics (e.g., ethnicity, gender, etc.).

The principles conclude with a note that they are deliberately underspecified in order to enable broad application in a wide variety of contexts. Because they had already been discussed in detail elsewhere, two important principles have been deliberately omitted. These concern data protection (with reference to the OECD Privacy Principles) and the impacts of human experiments (with reference to the Belmont Report).

The second section of the FAT/ML document consists of a collection of questions and steps for the preparation of a Social Impact Statement for Algorithms, which is aimed at ensuring compliance with the aforementioned principles. The creation of such a statement by the developers of algorithms should be carried out a total of three times during the process of design and development: during the design phase, pre-launch and post-launch. The Social Impact Statement should also be published after the launch of the algorithmic system to allow the general public to assess the system accordingly.

The structure of the Social Impact Statement is again based on the five principles described above. In each case, a number of key questions are specified and first steps are taken toward their achievement. For example, the first principle of accountability inquires as to which person or persons are responsible for damage to users, and the options that are available for reporting or recourse. Concrete steps for the clarification of these issues include the naming of an individual who is responsible for the social impact of the algorithm, the provision of contact information for complaints, and the development of plans for responding to unwanted side effects of the system.

Further measures from the Social Impact Statement include the following:

- *Explainability:* the creation of a plan that explains the decision-making process to users, allows data subjects to access and modify their data as needed, describes the training data used in any machine-learning models, and which explains the changes made to, and rectification of, underlying data.

- *Accuracy:* an assessment of the error and damage potential of the system; performance of a sensitivity analysis that provides insights into how the uncertainty of the output correlates with the uncertainty of the input; the development of a process that enables people to revise errors in input and training data, as well as in output decisions; performance of a validity test and subsequent publishing of the overall data error rate using the example of a random set.

- *Auditability:* the documentation and provision of an API that enables third parties to scrutinize and appraise the algorithmic system; measures to ensure that the terms of use permit the scientific community to carry out automated public audits; the creation of a plan for communication with third parties interested in auditing the algorithmic system.

- *Fairness:* the seeking of contact with experts familiar with the subtle social context that is formed by the system and in which the system is formed; assessments as to whether individual affiliation to specific categories (e.g., ethnicity, gender, gender identity, religion, socioeconomic status, etc.) negatively impacts the results of the algorithm.

5.2 Analysis

5.2.1 Strengths

Because the FAT/ML Principles concentrate exclusively on the aspect of accountability, a clear strength of the document is its more detailed understanding of this criterion than is the case in other quality criteria catalogs. In the demands laid out by the ACM U.S. Public Policy Council, accountability is referred to as a single aspect that – as in the Asilomar Principles – appears to be equated with responsibility. Thereby, one of the particular achievements of the FAT/ML Principles is that it adopts a far more differentiated concept of accountability, which is itself comprised of further criteria. Moreover, in their entirety, the five constitutive criteria for accountability aim at both technical (accountability, explainability, accuracy, auditability) and moral adequacy (fairness).

A further advantage of the document is the basic assumption (explicated in the premise) that the responsibility for the decisions of an algorithmic system is always located with humans. This is an indispensable foundation for professional ethics, above all against a backdrop of media descriptions of the impotence of human actors against machines, and the “algorithm is to blame” apology that is invoked all too often by companies and institutions.

The clear strength of the FAT/ML catalog is the practicability of the proposed quality criteria. The presence of a concrete draft of a Social Impact Statement clearly differentiates this document from previously analyzed proposals. Included is not only a description of such a statement, but also specific questions to be answered and important actions to be initiated. The creation of a Social Impact Statement is thus made considerably easier for the responsible parties. Here, it is worth noting that, more than other quality criteria compendia, the FAT/ML document successfully negotiates the often purely abstract subject matter.

In addition, the authors include a requirement for responsible parties to publish the Social Impact Statement, with the aim of facilitating a public discourse on the respective algorithmic system. As mentioned above, the role of the informed public should not be underestimated during the design of algorithmic systems, as this represents an available forum for a discussion that extends beyond economic interests. Nevertheless, to avoid industrial espionage, data breaches and manipulation of systems, it is often wise to avoid giving complete insights into the workings and data basis of software. A Social Impact Statement enables the public to assess social relevance and potential risk without allowing sensitive data to fall into the wrong hands.

Lastly, it can be said that the criteria of the FAT/ML catalog are directed not only toward the programming process, but also give meaningful consideration to downstream steps, such as appeals of incorrect decisions.

5.2.2 Weaknesses

Following on from the latter observation, it is regrettable that while the FAT/ML document does indeed consider downstream steps in the design of algorithmic processes, it does not explicitly deal with related upstream steps. For example, there is a complete lack of recourse to research into algorithmic processes and their moral integrity.

As likewise mentioned above, the authors of the FAT/ML catalog deal explicitly with the concept of accountability, which contributes to a more differentiated picture of this aspect. However, no consideration is given to other important quality criteria for algorithms, such as the security of the system or data protection. In particular, the moral requirements of the design of algorithmic processes are not sufficiently addressed. As already mentioned, of the five quality criteria, five are technical in nature and only the aspect of fairness addresses the ethical dimension of the design process.

That an algorithm is fair in its decisions (or preparation of such decisions) is without question an important requirement in socially sensitive contexts. However, fairness is merely a comparative concept, as it that implies that every decision must be made on the basis of the same set of parameters. For example, an individual must not be penalized for their skin color or gender if these criteria are not disadvantageous for others in the same context. It should nevertheless be noted that not every immoral decision is necessarily unfair – in principle, it may violate moral rules in equal measure for every individual. Unfortunately, the FAT/ML catalog overlooks this dimension. For example, there is a complete lack of reference to human rights, the moral values of a society, or the rule of law.

Finally, it can be observed that the FAT/ML statement rests on a somewhat narrow understanding of the circle of people that are responsible for algorithmic systems. The preamble is aimed explicitly at developers and product designers, while the Social Impact Statement is to be composed solely by “algorithm constructors.” Following on from Zweig (2018), it is possible to identify a large number of other actors in the social design of software systems. Contracting entities, policy makers, public institutions and implementers represent just a few of the groups that are unfortunately not addressed by the FAT/ML Principles.

In summary: The FAT/ML Principles are a set of quality criteria that applies a differentiated parameterization of the aspect of *Accountability,* and which discusses in detail its application in algorithmic systems. However, this tight focus also causes the neglect of other important criteria, and greater sensitivity to ethical (and not merely technical) issues in the social design of algorithms would be desirable.

The particular importance of the FAT/ML document is in its delineation of a Social Impact Statement. This provides concrete steps for the public documentation and appraisal of the social relevance of an algorithmic system. As such, this collection of guidelines for action circumvents the issue of an overly general formulation of abstract criteria, as is characteristic of many other compendia of quality criteria.

6 Conclusions

In the following section, we will use the analysis of the above documents to draw conclusions for the creation of a catalog of quality criteria within the German-speaking context. In doing so, we will focus on the following issues:

a) *Transferable strengths:* In creating a catalog of quality criteria for the design of socially relevant algorithmic processes, what assumptions, content and format-related elements of other such collections would be useful to adopt? For each aspect discussed, we will also identify the document in which it takes exemplary form. The strengths will be considered within the following categories:

   - *Formal traits*
   - *Content-related traits*
   - *Specifications on implementation*

b) *Correctable weaknesses:* What important aspects were not considered in the analyzed documents?

6.1 Transferable strengths

An important basic premise – that the responsibility for algorithmic processes must be exclusively situated with human actors – should be included at the level of presuppositions in order to formally justify the creation of the catalog of quality criteria. Foregrounding this premise in a preamble helps clarify the rationale behind a collection of quality criteria that is aimed at human designers of algorithmic processes. In addition, this premise creates a counterpoint to the scenario, today found widely in the media, of humans' powerlessness when confronted with machines; moreover, it forestalls any possible denial of responsibility on the part of those involved in the design process. A well-formulated one such premise can be found in the FAT/ML principles.

As a further premise, it is important to outline the social significance of algorithmic processes, or at least to formulate this as a basic assumption. Emphasizing the relevance of software systems for social inclusion and participation helps establish the significance of a code of ethics for their design; moreover, this can provide actors with motivation to comply with the criteria. This premise is contained in all three documents. However, they differ in their approach, at times very significantly. The ACM U.S. Public Policy Council's paper, like the FAT/ML principles, contains only a brief description of the pervasiveness of algorithmic systems, although the latter document also makes reference to the "potential for significant societal impact" (FAT/ML 2018). The Future of Life Institute's Asilomar AI Principles take a different approach. Because this document's substantive focus is on artificially intelligent systems, it makes reference to AI's far-reaching and potentially disastrous influence on society. In addition, it notes that no strong assumptions regarding the upper limits of the capacities of artificial intelligence should be made, because there is as yet too little firmly established knowledge on this subject. In contrast to the two other documents, the Asilomar AI Principles do not include these statements in a preamble, but rather as individual principles in the *Longer-Term Issues* section. However, the introduction contains the optimistic assumption that, "guided by the following principles, [AI] will offer amazing opportunities to help and empower people in the decades and centuries ahead." (FAT/ML 2018). Whether a document's drafters take a positive, neutral or negative basic attitude toward the social significance of algorithmic processes, is evidently a contentual issue. Reference to this stance should nevertheless be placed before the actual catalog of quality criteria, preferably in a preface, in order to render the authors' motivation as clear as possible.

With regard to content-related demands, a series of strengths can be derived from the documents analyzed above.

First of all, it is important to do justice to the complexity of the algorithmic design process. Rather than focusing solely on the programming step, the quality criteria should also include upstream and downstream steps. For example, the documents analyzed above incorporate the specification of goals for a system’s use, the collection of data, the process of training an artificially intelligent system, and the ability to reconstruct, understand and challenge failures. A closely associated insight is that sources of error may not necessarily be located in the system itself but can also stem from the data being employed. The ACM U.S. Public Policy Council’s paper and the FAT/ML principles both address data-quality requirements that derive from this fact.

As the diversity of the algorithmic-system design process is reflected in the catalog of quality criteria, the document’s target audiences inevitably expand. Capturing the entirety of the design process is possible only by addressing programmers as well as politicians, customers and clients, institutions and, to some extent, even users. In the Asilomar AI Principles this task is performed particularly well. When drawing up a catalog of quality criteria for a German-language context, it will be beneficial to include as many relevant actors from different areas as possible in the process in order to arrive at a representative document.

In this regard, two groups of actors – the research community and the public – require special attention. Researchers in the area of algorithmic decision-making and artificial intelligence are ultimately responsible for what will prove technically feasible in the future. Researchers are explicitly taken into account only in the Asilomar AI Principles, which calls for them to develop “beneficial intelligence” (Future of Life Institute 2017a). This sensitivity to the key role played by researchers is exemplary and should be regarded as a standard to be followed within any future quality-criteria projects. Even researchers working in areas outside computer science can make significant contributions to the design of algorithmic processes. For this reason, the Asilomar AI Principles’ call for the provision of research funding to relevant social-science, economics, ethics and law projects is highly sensible.

The public’s role as an additional important actor is addressed only by the FAT/ML principles and by the ACM U.S. Public Policy Council. Promoting public discourse regarding an algorithmic system’s participatory relevance or appropriateness is vital, particularly because this format provides a discussion platform for those who will (presumably) be affected. In addition, the publication of social impact statements (as called for by the FAT/ML principles) can help prevent the emergence of competitive races between rival providers at the cost of safety (as noted in the Asilomar AI Principles), while keeping the public welfare in view as an overarching goal of algorithmic processes. As the ACM U.S. Public Policy Council notes, an additional advantage of integrating the public into the process of designing algorithmic systems lies in the enormous increase in the number of people reviewing a system for failures. However, this does not mean that one should rely on failures being detected only when a system has entered productive operation. The far-reaching consequences for those affected are much too high for such a trial and error approach (see Rohde 2017 on the robo-debt scandal in Australia).

Taking account of the multiplicity of actors involved in designing algorithmic processes inevitably runs the risk of exposing conflicts of interest between them. If the document is to be credible, this issue must not be glossed over in an effort to preserve the catalog’s simplicity. Of course, it is impossible to describe all possible scenarios fully. For this reason, a reference to the problem, following the ACM U.S. Public Policy Council’s example, can be useful. In the introduction of that document, it is noted that technical, economic and social interests can render the software-design process opaque; moreover, this is illustrated using short examples. One frequently cited example mentioned in the FAT/ML principles, and which serves as sufficient acknowledgment of the problem, is the difficulty of balancing full system transparency with protections for data privacy and trade secrets.

A catalog of quality criteria for the use of algorithms in socially sensitive situations must be more than a simple collection of technical quality characteristics such as a system’s auditability quotient or the degree to which its failure rate has been reduced. Indeed, an algorithm's moral quality must also be taken into account. This includes comparative concepts such as fairness, which is described in detailed form in the FAT/ML principles, but also more fundamental ethical questions. The Asilomar AI Principles are exemplary in this respect, as they include respect for human rights, among other issues. Indispensable too is the compatibility of algorithmic systems with further human values such as dignity, rights, freedoms and cultural diversity, as also described in this document. The Asilomar AI Principles additionally employ an ideal approach to meta-issues associated with the configuration and fundamental goals of an artificially intelligent system.

Finally, having a detailed understanding of the individual quality criteria is beneficial. Because recurring buzzwords such as accountability, transparency and auditability are frequently used in discussions of the ethics of algorithms, it is critical for target audiences that these terms be adequately defined. To be sure, the precision with which the FAT/ML principles treat the aspect of accountability, for example, is not necessarily the right strategy for all characteristics. However, the distinction between *Judicial Transparency* and *Failure Transparency* (Asilomar AI Principles) provides a helpful example of the conceptual nuances among individual criteria. In short: a detailed conceptualization of individual criteria is certainly desirable; however, in the interests of user-friendliness, this should be reduced to short definitions.

With regard to the *formal design* of a collection of quality criteria, list formats offer the basic advantage of readability. Indeed, all three catalogs analyzed here take this into account. Moreover, there is the question of how best to ensure that a collection of quality criteria can in fact be applied. The above analysis offers two successful strategies: giving the catalog of demands a legislative-text-like nature, or compiling a collection of concrete steps. The Asilomar AI Principles in particular take the first path. This document's principles are formulated both generally and with an idealistic orientation toward ambitious goals, leaving room for interpretation in individual cases. As previously noted, this offers a key advantage; the quality criteria can reflect the diversity in the overall process of algorithm design, without simply describing special fields of use.

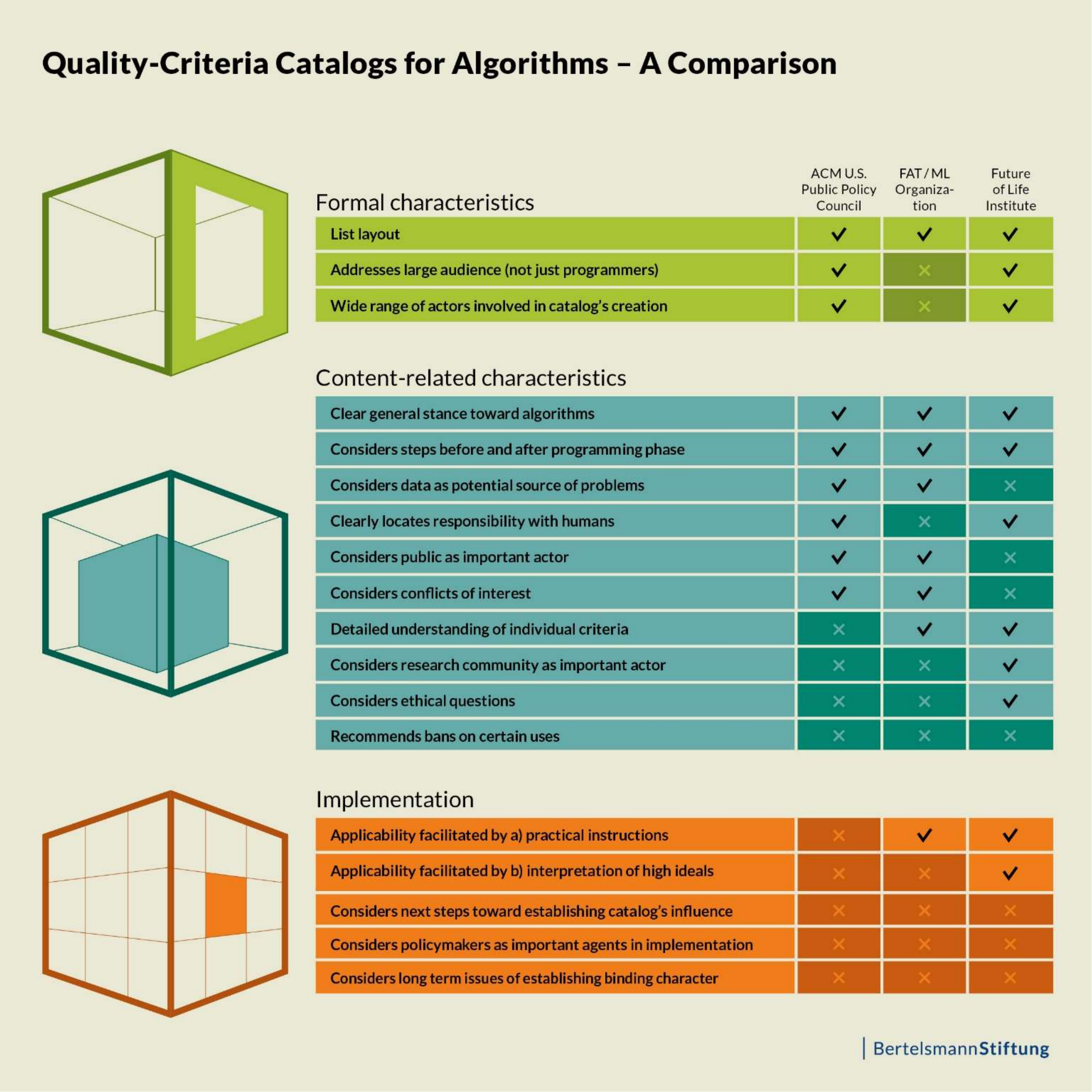
A second possibility, represented here particularly by the FAT/ML principles' Social Impact Statement, is to ensure applicability through the specification of concrete steps. To some extent, this is also realized by the Asilomar AI Principles section on research funding. The formulation of concrete first steps for complying with or demonstrating compliance with quality criteria facilitates implementation by those responsible, creates incentives and enhances the impression that the project is, in fact, feasible. Ultimately, we cannot here resolve the question as to whether those creating collections of quality criteria should employ legislation-like text or compose a collection of prescriptive guidelines. Indeed, authors should consider combining the two approaches.

6.2 Weaknesses to be corrected

All three collections of quality criteria analyzed above are alike in offering no specific pointers for practical implementation of the document – specifically with regard to making it a binding standard. Their practical recommendations relate at best to the implementation of individual criteria, not to how the collections of quality criteria as a whole can become a new standard. Strictly speaking, such questions of implementation are not part of the quality criteria. However, the lack of reference to this issue means an important stimulus for the ethical design of algorithmic processes is missing. Also the role of politics in this context is largely left unacknowledged in all three of the documents analyzed.

There has been little academic research into what makes successful codes of professional ethics in other professional fields binding. This can explain the absence of such references in the existing documents. Within the Bertelsmann Stiftung's *Ethics of Algorithms* project, this question is currently being addressed by Dr. Alexander Filipović's team of authors. The findings resulting from this expert research will be published in the coming weeks and should be integrated into a catalog of quality criteria for the German-speaking context. In any case, a short paragraph recommending concrete steps for implementing the document should be included, and should perhaps even be tailored to certain groups of responsible parties such as policymakers, public institutions and developers.

The issue of prohibitions is closely linked to the above question of bindingness. None of the above-cited compendia of quality criteria consider methods of prohibiting particular algorithmic systems in cases of non-compliance with particular demands. Indeed, the neglect of this issue is a common phenomenon in the debate over algorithmic decision-making and artificial intelligence (see Krüger and Lischka 2018). One goal of a catalog of quality criteria should be to prevent the use of systems that fail to meet the described requirements. The document must make this obvious, because the absence of any reference to possible penalties leaves the issue of sanctions open, and thus also the degree to which the quality criteria are in fact binding. Particularly where the *ethical* adequacy of a software system cannot be guaranteed (for example, when there is a risk that society's moral values or human rights may be subverted), a categorical ban on use in a production operation should be possible to declare.



Quality-Criteria Catalogs for Algorithms – A Comparison

Formal characteristics	ACM U.S. Public Policy Council	FAT / ML Organiza-tion	Future of Life Institute
List layout	✓	✓	✓
Addresses large audience (not just programmers)	✓	✗	✓
Wide range of actors involved in catalog's creation	✓	✗	✓
Content-related characteristics			
Clear general stance toward algorithms	✓	✓	✓
Considers steps before and after programming phase	✓	✓	✓
Considers data as potential source of problems	✓	✓	✗
Clearly locates responsibility with humans	✓	✗	✓
Considers public as important actor	✓	✓	✗
Considers conflicts of interest	✓	✓	✗
Detailed understanding of individual criteria	✗	✓	✓
Considers research community as important actor	✗	✗	✓
Considers ethical questions	✗	✗	✓
Recommends bans on certain uses	✗	✗	✗
Implementation			
Applicability facilitated by a) practical instructions	✗	✓	✓
Applicability facilitated by b) interpretation of high ideals	✗	✗	✓
Considers next steps toward establishing catalog's influence	✗	✗	✗
Considers policymakers as important agents in implementation	✗	✗	✗
Considers long term issues of establishing binding character	✗	✗	✗

BertelsmannStiftung

7 Sources

Association for Computing Machinery (ACM) (2018). “ACM U.S. Public Policy Council.” https://www.acm.org/public-policy/usacm (Download 14.3.2018).

Association for Computing Machinery (ACM) (2017). “Statement on Algorithmic Transparency and Accountability and Principles for Algorithmic Transparency and Accountability.” www.acm.org/binaries/content/assets/public-policy/2017_joint_statement_algorithms.pdf (Download 14.3.2018).

Fairness, Accountability, and Transparency in Machine Learning (2018). “Principles for Accountable Algorithms and a Social Impact Statement for Algorithms.” www.fatml.org/resources/principles-for-accountable-algorithms (Download 14.3.2018).

Fairness, Accountability, and Transparency in Machine Learning (FAT/ML) (2016). “Dagstuhl Seminar 16291. Data, Responsibly.” www.dagstuhl.de/en/program/calendar/semhp/?semnr=16291 (Download 14.3.2018).

Filipović, Alexander, Claudia Paganini and Christopher Koska (2018). “Verbindlichkeit erfolgreicher Professionsethiken und Übertragbarkeit auf gesellschaftlich relevante algorithmische Prozesse.” Forthcoming. Eds. Bertelsmann Stiftung. Gütersloh.

Future of Life Institute (2018a). “Existential Risk.” https://futureoflife.org/background/existential-risk/. (Download 14.3.2018).

Future of Life Institute (2018b). “The Future of Life Institute (FLI).” https://futureoflife.org/team/ (Download 14.3.2018).

Future of Life Institute (2017a). “Asilomar AI Principles.” https://futureoflife.org/ai-principles/ (Download 14.03.2018).

Future of Life Institute (2017b). “Beneficial AI 2017.” https://futureoflife.org/bai-2017/ Download (14.3.2018).

Future of Life Institute (2017c). “A Principled AI Discussion in Asilomar.” https://futureoflife.org/2017/01/17/principled-ai-discussion-asilomar/ (Download 14.3.2018).

Krüger, Julia, and Konrad Lischka (2018). “Damit Maschinen den Menschen dienen. Lösungsansätze, um algorithmische Entscheidungen in den Dienst der Gesellschaft zu stellen.” Forthcoming. Eds. Bertelsmann Stiftung. Gütersloh.

Rohde, Noëlle (2017). “In Australien prüft eine Software die Sozialbezüge – und erfindet Schulden für 20.000 Menschen.” *Algorithmenethik.de* 25.10. https://algorithmenethik.de/2017/10/25/in-australien-prueft-eine-software-die-sozialbezuege-und-erfindet-schulden-fuer-20-000-menschen/ (Download 15.3.2018).

Zweig, Katharina (2018): *Wo Maschinen irren können. Fehlerquellen und Verantwortlichkeiten in Prozessen algorithmischer Entscheidungsfindung*. Eds. Bertelsmann Stiftung. Gütersloh.

8 About the author

Noëlle Rohde works as a Project Manager in the team "Ethics of Algorithms" at the Bertelsmann Stiftung and is responsible for the "Professional Ethics" module. She studied Medical Anthropology at the University of Oxford, where she focused her dissertation on the concept of quantificational discrimination in Global Health metrics and self-tracking techniques. She previously studied Philosophy, Psychology and Linguistics at the Universities of Paderborn and Oxford.

Address | Contact

Bertelsmann Stiftung
Carl-Bertelsmann-Straße 256
33311 Gütersloh
Telefon +49 5241 81-0

Ralph Müller-Eiselt
Senior Expert Taskforce Digitalization
Telephone +49 5241 81-81456
ralph.müller-eiselt@bertelsmann-stiftung.de

Noëlle Rohde
Project Manager
Telephone +49 5241 81-81141
noelle.rohde@bertelsmann-stiftung.de

www.bertelsmann-stiftung.de

BertelsmannStiftung