UNESCO Recommendation on the Ethics of Artificial Intelligence

Conditions for the Implementation in Germany
UNESCO Recommendation on the Ethics of Artificial Intelligence. Conditions for the Implementation in Germany

Prof Dr Matthias C. Kettemann

II. Added Value of the UNESCO Recommendation on the Ethics of AI Compared to Other International Initiatives

   II.1. OECD Recommendation on Artificial Intelligence
   II.2. Council of Europe: CAHAI
   II.3. Draft Regulation of the European Union
   II.4. UNESCO Recommendation on the Ethics of AI: Approach and Added Value
   II.5. Summary

III. Conditions for the Implementation of the UNESCO Recommendation on the Ethics of AI in Germany

   III.1. Introduction
   III.2. Ethical Impact Assessment (Policy Area 1)
   III.3. Ethical Governance and Stewardship (Policy Area 2)
   III.4. Data Policy (Policy Area 3)
   III.5. Development and International Cooperation (Policy Area 4)
   III.6. Gender (Policy Area 6)

IV. Conclusion

Imprint
“Even though much is still open, one thing is clear: we as humanity still have the opportunity to guide the future development of AI and to ethically frame it in a humanistic, human rights perspective. UNESCO has taken up this issue as part of its mandate.”
— Prof Dr Maria Böhmer, President of the German Commission for UNESCO
Artificial intelligence (AI) has found its way into many areas of our lives in recent years and fundamentally changed them. It provides personalised search results on the World Wide Web, facial recognition on smartphones and semi-automated driving. The use of self-learning algorithms is a megatrend of our time, but at the same time the long-term consequences of the use of AI can hardly be predicted. Thus, the current discourse on AI and its effects spans a wide range, from utopias of a better and fairer world to dystopias of an uncontrollable superintelligence that dominates humans. Even though much is still open, one thing is clear: we as humankind still have the chance to steer the future development of AI and to ethically frame it within humanistic and human rights-based boundaries.

UNESCO has taken up this issue as part of its mandate. After a long negotiation process involving all 193 Member States and a wide range of stakeholders, UNESCO adopted a landmark Recommendation on the Ethics of Artificial Intelligence in November 2021. The UNESCO Recommendation is the first globally negotiated response to one of the greatest challenges of our time. It convinces with its strong human rights language, a clear ethical foundation and concrete tasks for action for policymakers. Thus, it fulfills all requirements to become a global compass for responsible AI development and use.

The Recommendation defines ethical guidelines as well as concrete tasks for the implementation in policy areas in which AI is already widely discussed today. Beyond that, the inclusive negotiation process has brought areas into focus that have so far played a rather minor role in the AI discourse, but are at the core of UNESCO’s mandate. The UNESCO Recommendation contains tasks for action in eleven different policy areas, including education and science, communication, health and the environment. Noteworthy is the Recommendation’s focus on previous “blind spots” of AI regulation, such as AI and gender, AI and sustainability, AI and the Global South, and AI and education. In addition, it specifies clear procedural and transparency obligations, such as a multi-stakeholder-based, open and responsive design of relevant processes.

The German Commission for UNESCO, as a link between the UNESCO Secretariat, national politics and civil society, supports the implementation of this important international law document in Germany. This study serves as a substantive foundation for this task. It takes a look at the status quo and clarifies which processes, instruments and laws already exist in Germany that support AI development in accordance with the recommendation.

At the same time, it also analyses where regulation or action is still needed and offers concrete proposals for the implementation of the UNESCO Recommendation in Germany. In addition, it elaborates on the added value of the Recommendation in comparison to other initiatives, such as the OECD Recommendation or the European Commission’s proposed Regulation. Just two examples for this added value: The UNESCO Recommendation is the only legal document to refer to cultural aspects that may be affected by the use of AI systems. Furthermore, it strongly emphasises the need to protect marginalised groups in the context of AI development.

We would like to extend our sincerest thanks to the author of the study, Professor Matthias C. Kettemann. We need every effort from politics and administration to ensure that AI development proceeds in an ethical and human rights-based manner. AI dystopias must remain science fiction!
SUMMARY AND POLICY RECOMMENDATIONS

SUMMARY: THE ADDED VALUE OF THE UNESCO RECOMMENDATION ON THE ETHICS OF AI

→ The UNESCO AI Ethics Recommendation is the first international legal text on AI ethics that has a global reach and was negotiated in an open and inclusive multi-stakeholder process. It includes globally accepted ethical standards for AI-based technologies and applications in full respect of international law, in particular human rights. Thus, human rights can now play a central role in the development of AI-related standards around the world. 193 Member States have committed to the Recommendation. The open process over two years, in which non-state actors were also able to contribute their views, represents a significant added value in terms of the Recommendation’s legitimacy.

→ Remarkable in terms of content is the Recommendation’s holistic approach. At a general level, ethical AI development and use is defined as a central challenge and obligation for the whole of society. On the level of AI governance, the Recommendation formulates concrete tasks for action in eleven policy areas, including education and science, communication, health and the environment.

→ The focus on ethical rules, which are, however, closely linked with and justified by human rights, is deliberately chosen. Clearly tailored ethical rules and principles relate to law in various ways; they can help in the development and implementation of policies and in the interpretation of legal norms by providing orientation points for technological development.

→ Noteworthy is also the Recommendation’s focus on normative “blind spots” in other areas, such as AI and gender, AI and sustainability, AI and the Global South (development), AI and education, as well as on clear procedural and transparency obligations: Specifically, it calls for ethical impact assessments and a multi-stakeholder-based, open and responsive design of the corresponding processes.
Policy Recommendations for AI Actors

1. **AI must respect, protect and promote human rights**
   Worldwide, and especially in Germany, every decision on AI – as a contingent technology – must be aligned with ethical values, human rights and international law in the process of planning and deployment.

2. **Ethical (and legal) guidelines must be followed throughout the whole life cycle of AI systems**
   AI is no deus ex machina. Therefore, all AI stakeholders must take the entire life cycle of AI systems into account – a process that begins with the formulation of the requirements for an AI system and the composition of the development team that decides on training data and the conditions for machine learning.

3. **AI data must be as open to the public as possible**
   The German government must ensure that the scientific community has access to AI research data in a privacy-conscious manner and that this data is not monopolised by large companies. In addition to the scientific community, this access must also be guaranteed for selected civil society actors who follow different logics and interests than the scientific community and can assume a public watchdog function.

4. **AI must contribute to a diverse society**
   Gender, inclusiveness and diversity mainstreaming throughout the entire life cycle of AI applications are crucial to prevent risks of discrimination. In order to ensure that the content of training data is free of stereotypes and bias, documentation obligations must be created for AI stakeholders in Germany. Developers and research teams must also be more diverse. But non-discrimination and equal treatment are systemic issues: politics and administration in Germany should therefore take measures in all policy areas to actively promote equity and equality for women and diverse people.

5. **AI must be publicly monitored**
   The German government must develop and establish clearer frameworks for the implementation of ethical impact assessments and monitoring mechanisms. These must include a right to disclosure of transparency protocols for users. Only internal monitoring mechanisms are not sufficient. Corporate AI stakeholders must ensure that the results of AI decision-making processes are comprehensible, explainable and justifiable. Human beings must not feel at the mercy of AI. On the contrary: All AI actors are obliged to increasingly introduce citizens to the role and potential of AI by means of modern knowledge formats.

6. **AI must become globally accessible**
   AI stakeholders, especially the German government, should help build AI capacity in Low- and Middle-Income Countries (LMICs) and Least Developed Countries (LDCs) to correct contextual biases of ‘Western AI’. It is precisely the lack of access to big data, which is crucial in AI development, that will deepen existing gaps unless deliberate countermeasures are taken. In addition to public development cooperation, the role of the private sector is particularly important. The private sector should act as a partner and should increasingly be held accountable by the federal government.

7. **AI must contribute to the realisation of the SDGs**
   Digitisation must be sustainable, and sustainability must be digitally designed. AI must always be assessed with a view to promoting the achievement of the Sustainable Development Goals (SDGs). In general, the use of AI must (also) be classified in terms of ecological sustainability.

8. **AI standards must be negotiated in inclusive processes**
   The German government and all involved AI companies and standardisation bodies must enable greater civil society engagement in standardisation processes as industry interests are traditionally over-represented here.

As many AI stakeholders as possible should be involved, especially representatives of civil society, data protection authorities and consumer protection organisations as well as representatives of vulnerable groups.
I. Introduction

The UNESCO Recommendation on the Ethics of Artificial Intelligence is a contribution to a global approach to digital human rights. For too long the protection of human rights in digitalization has been sought within national normative logics by too many states. The German Federal Government’s 14th report on its human rights policy includes the strengthening of human rights protection “in the face of digital transformation” as a priority in the “Human Rights Action Plan” for the years 2021–2022. This also includes the protection of human rights in the context of the development of Artificial intelligence (AI).

Given the diversity of economic, social, cultural, civil and political rights whose respect, protection and enforcement are affected by AI, there is a need for principles and rules to realise the potential of AI while protecting individual freedoms and guaranteeing social cohesion. Against this background, UNESCO has developed its Recommendation on the Ethics of Artificial Intelligence (hereinafter referred to as: “Recommendation”).

The German Government’s Coalition Agreement 2021–2025 does not mention human rights in the context of AI, while referring to AI as a “future sphere”. At the same time, the use of the potentials of AI is stipulated and the importance of an intensive transatlantic dialogue on data sovereignty, web freedom and AI is emphasised. This shows that a debate on the human rights and ethical aspects of AI development and use is crucial. Building on this, the definition of clear ethical principles and operationalisable tasks for relevant groups of actors is necessary. In this context, the UNESCO Recommendation on the Ethics of Artificial Intelligence is a contribution to a globally conceived, common digital human rights policy.

The Recommendation, adopted on 23 November 2021, provides an “ethical guiding compass and a global normative bedrock allowing to build strong respect for the rule of law in the digital world.” This study focusses on what this can mean for Germany.

First, the study looks at the added value of the UNESCO Recommendation in comparison to other international initiatives. This is necessary against the background of the considerable number of 173 sets of rules on AI ethics contained in AlgorithmWatch’s “AI Ethics Guidelines Global Inventory”, which was updated in mid-2020.

Principles must become relevant in practice. Therefore, the study analyses – with a view to specific policy areas – the conditions for the implementation of the Recommendation in Germany. An important output of the study is a summary with policy recommendations for political practitioners and multipliers.

“The UNESCO Recommendation on the Ethics of Artificial Intelligence is a contribution to a global digital human rights policy, which unfortunately is still thought of only nationally in too many states.”

---

4 The heavily used term “ethics” in the current technology debate also requires reflection. What is important is the independent consideration of non-legal normative systems, the added value that lies in their formal non-binding nature and in their significance for the flexibility in dealing with changing technology. However, it is important to point out the undeniable advantages of legal systems of order (which can be enforced by coercive force, if necessary).
Characteristics

The Recommendation is the first globally negotiated text under international law in the field of AI ethics. It is not only global, but also holistic in terms of its approach. The Recommendation offers the 193 UNESCO Member States a framework for action in this important future field (by contrast, the Council of Europe’s AI rules to be discussed only apply to 47 states; 46 states were involved in the OECD AI standards). UNESCO Member States include not only states with a solid rule of law, such as Germany, in which – via the European level – AI standards are already introduced or currently negotiated. They also encompass states such as China, which produces the lion’s share of AI and use it in a challenging way from a human rights perspective. In addition, there are also those Member States that have neither fundamental rights protection against AI nor national AI capacities. Observers found it particularly noteworthy that China, which was critical of certain human rights provisions in the Recommendation during the negotiations, ultimately supported the consensus.1 Although the largest user of AI (and home to most AI-based start-ups), the USA, is not a UNESCO Member State at the moment, it is known that the USA followed the process of drafting the Recommendation closely within the scope of their observer status.

Process

The UNESCO Recommendation was developed over two years in an intensive and sometimes controversial intergovernmental negotiation process. This multi-stakeholder approach, which was enriched by several regional stakeholder consultations, already offers significant added value in terms of the Recommendation’s legitimacy. By involving various stakeholder groups, the Recommendation thus followed a “best practice” model for international standard-setting.

Content

With specific reference to eleven policy areas, including education and science, communication, health and the environment, the Recommendation translates principles for a dignity-sensitive use of AI into policy-making tasks.

The aim of the Recommendation is to align AI with human rights. It gives AI an ethical foundation that not only protects human rights and human dignity, but addresses all three dimensions of human rights protection: respect, protection and promotion.

Ethics and Rights

The focus on ethical rules, and not primarily on human rights obligations, is deliberately chosen. On the one hand, UNESCO has a special responsibility with regard to ethical considerations in important social agendas and discourses. On the other hand, clearly defined ethical values and principles relate to law in various ways; they can support the development and implementation of policies and the interpretation of legal norms by, as the Recommendation puts it, “providing guidance with a view to the fast pace of technological development”. Ethics is not “lesser” than human rights, but an alternative (albeit congruent in terms of protective intent in many areas). Ethical rules are conceived and structured differently. They are not centrally controlled, and enforced in another way than human rights, regularly not by coercive force. States that enter into human rights obligations are bound by them under international law. States that commit themselves to ethical obligations can only – but at least – be induced by international pressure to behave in accordance with their obligations. While for some states the obligation arising from human rights treaties is limited to signing them, it is obvious that ethical commitments have to be fulfilled through “striving efforts” by states. Especially in a fairly new field such as AI, legal rules have not yet reached consensus – especially internationally. Ethical rules have an important function here.

Aim

The Recommendation aims to define globally accepted ethical standards for AI technologies in full respect of international law, in particular human rights, which can then be key to the development of AI-related norms around the world.

The Recommendation is further characterised by a holistic focus on the different policy fields, an awareness that each field requires different regulatory approaches and the focus on “blind spots” of previous AI regulations. These include environmental protection, the sustainable and resource-efficient use of AI, and the use of AI in education, with full recognition of the right to education for all.

Normative technique

Depending on the field of action, the recommended measures have a different degree of obligation. Although the Recommendation as a whole is uniformly “soft law”, i.e. not formally binding law, it was negotiated in such a detailed manner that its quality differs largely from simple resolutions or declarations. The Recommendation cannot be enforced in courts, but it becomes effective towards states. Among other things, it calls for concrete governance measures, e.g. an Ethical Impact Assessment (EIA) for AI systems or a network of independent AI ethics officers to monitor the EIA. Beyond this, the Recommendation calls for international cooperation and research in the field of AI, and a reconsideration of the usually means-driven selection of measures by states to ensure ethical AI use.

---

Added Value of the UNESCO Recommendation on the Ethics of AI Compared to Other International Initiatives
II. Added value of the UNESCO Recommendation on the Ethics of AI compared to other international initiatives

In order to make a sound assessment of the added value of the UNESCO Recommendation, it is necessary to first look at other international ethics-based AI initiatives. This shows in which policy and governance areas the UNESCO Recommendation makes new, further-reaching proposals, restructures existing ones or introduces new topics, instruments or approaches to governance. Furthermore, it becomes clear to what extent the Recommendation also provides structural added value besides content-related aspects.

In addition to the UNESCO Recommendation, other AI initiatives of particular relevance to Germany come from the Council of Europe and its Ad hoc Committee on Artificial Intelligence (CAHAI), the European Union and the OECD. They all have issued guidelines and laws on the ethical use of AI or are currently in the process of negotiating them.

II.1. OECD Recommendation on Artificial Intelligence

The Organisation for Economic Cooperation and Development (OECD) has recognised the need to develop international policy instruments in the context of the widespread development and deployment of AI systems. In the course of this, the OECD Committee on Digital Economy Policy (OECD/CP), with the involvement of governments, industry, civil society and trade unions in an expert group (OECD Expert Group on Artificial Intelligence, AGOI) developed recommendations, which ultimately resulted in the adoption as the "Recommendation of the OECD Council on Artificial Intelligence" on 22 May 2019. 9

The OECD Recommendation begins by defining some terms [I.]: The definition of AI systems is so broad, however, that the question arises as to what the difference between AI systems and other software systems should be. The term "machine-based system" 10 does not help much either, as it is noteworthy here that the life cycle of an AI system, including its phases, is explicitly defined. In particular the phase of design, data collection and processing and modelling of the system is described as a "context-dependent sequence." 11 After in its own reflection first interventional standardisation initiative on AI, the OECD Recommendation sets out the following principles for the responsible governance of trustworthy AI in the first section: inclusive growth, sustainable development and quality of life as the goal of the use of AI systems [V.I.1.], human-centred values and fairness [V.I.2.], transparency and explainability [V.I.3.], robustness and security of systems [V.I.4.], and accountability of "AI actors" [V.I.5.].

In its second section on national measures and international cooperation, the OECD Recommendation provides for the promotion of a "digital ecosystem", which should include the digital technologies, infrastructures and mechanisms such as a data trust model necessary for an adequate exchange of "AI knowledge" [V.V.2.]. This exemplifies the adoption of a rather business-oriented perspective in the document. This impression is reinforced by the call for an "enabling policy environment for AI" [V.V.2.3.]. A quick transition should be achieved from the research and development phase to the implementation and operation of trustworthy AI systems [V.V.2.3.6].

Furthermore, countries are encouraged to review and, if necessary, adapt their policies and regulatory frameworks to foster AI-related innovation and competition [V.V.2.3.6-3]. Thus, a more economy-oriented position in favour of the use of AI principle cannot be dismissed in the OECD Recommendation.

II.2. Council of Europe: CAHAI

The Council of Europe is currently working intensively on Artificial Intelligence and its regulation in its bodies. On 11 September 2019, the Committee of Ministers mandated an Ad Hoc Committee on Artificial Intelligence (CAHAI) to conduct a study on the basis of multi-stakeholder consultations and the Council of Europe's human rights standards. This study was to determine the feasibility of a legal framework and its possible elements. 12 In December 2020, this study was finally published, 13 supplemented by a publication summarising the stakeholder perspectives contributed so far. 14

Furthermore, the OECD Recommendation assumes that an increasing implementation of AI systems in the economy, corresponding to a change in the world of work and society is necessary, so that humans must acquire the skills needed for the effective use of and interaction with AI systems [V.V.2.4.]. However, such assessments are rightly viewed critically, as they could possibly obscure other causes of social and economic problems as well as related countermeasures. 15 Overall, the OECD Council's Recommendation is a concise document that draws first lines for the more economy-oriented perspective. On the one hand, the values listed there are of such a universal and obvious nature that most stakeholders would probably not resist agreeing to them. On the other hand, they remain on an abstract level, so that the question of concrete forms, policy and governance measures remains open. This is a significant difference to the UNESCO Recommendation, which does not, but instead sets out normatively graded, policy field-specific demands at various points, which put different societal AI actors under obligation.

The process initiated in 2019/2020 continues to progress. Currently, the Legal Frameworks Group of CAHAI (CAHAI-LFG) has, after multi-stakeholder consultations and plenary meetings of CAHAI 16 developed a first feasibility study of the possible elements of legally binding instruments in the context of AI, human rights, democracy and the rule of law, which is now to be reviewed by CAHAI and subsequently submitted to the Committee of Ministers. 17

AI technologies, here characterised as services and products, are seen as having the potential to promote welfare as well as individual and societal prosperity. On the other hand, potential negative effects and dangers, especially with regard to the rights protected by the European Convention on Human Rights (ECHR), are also recognised [marginal no. 22]. 18 Examples of human rights risks identified in the study include AI systems that undermine the right to equality and non-discrimination by perpetuating prejudices and stereotypes (e.g. in employment), and AI-driven surveillance and tracking applications that threaten individuals’ rights to freedom of assembly and expression. 19 Furthermore, CAHAI explicitly recognises the lack of a uniform, universally recognised definition of the term Artificial intelligence, and classifies it as a collective term that should be approached in a technology-neutral manner [marginal no. 8.1-20].

The associated CAHAI feasibility study considers a wide range of instruments for regulating AI, including international legal instruments such as the ECHR and the EU Charter of Fundamental Rights [marginal no. 68 ff.], ethical guidelines for AI, including those developed by private companies and public sector organisations [marginal no. 74 ff.], and national instruments and policies for AI [marginal no. 76 ff.].

16 Ibid.
23 Ibid., p. 20 ff.
24 Ibid., p. 21 ff.
25 Ibid., p. 21 ff.
26 Ibid., p. 21 ff.
CAHAI’s feasibility study weighs up the advantages and disadvantages of these measures and finds that there is currently no binding international legal instrument specifically tailored to the challenges posed by AI systems (marginal no. 83 ff.). Instead, the many already existing instruments that party negotiators in each of them can use, to ensure that they are tailored to the AI-specific needs and peculiarities.

In response to the challenges, the CAHAI study sets out the core elements of a legal framework in the form of principles, rights and obligations (marginal no. 96 ff.). In addition to these key elements, two overarching measures aim to ensure that human rights are taken into account in the development and use of AI. Firstly, by ensuring that those affected can invoke their human rights in the AI context. Secondly, by describing the requirements that developers and users of AI systems must meet in order to protect human rights. CAHAI identifies seven essential principles to ensure the integration and protection of human rights in and by AI, including human dignity (marginal no. 98 ff.), the requirement of non-discrimination, gender justice, fairness and diversity (marginal no. 105 ff.).

The Recommendation of the CAHAI study to the Council of Europe represents a legally binding instrument which, based on the above principles, could serve as a basis for national legislation which should follow a risk-based and -differentiated approach (marginal no. 95 ff.), in addition, further soft and hard law instruments are recommended, which should be precise and tailored to the AI sector-specific needs and peculiarities.

While the CAHAI study is an expert study, the UNESCO Recommendation is the result of an intergovernmental negotiation process between UNESCO Member States. The CAHAI study therefore did not have to seek consensus. It is also noticeable that the study deliberately relates to the necessary development of legal regulations. With the elaboration of the AI-related human rights acquis the study is highly salient, albeit different from the policy field-specific Recommendation of UNESCO with its broad ethically based approach.

II.3. Draft Regulation of the European Union

The European Commission’s legislative proposal for AI regulation can be described as ambitious and will be binding to EU Member States after its adoption. On 21 April 2021, the European Commission presented the Draft Regulation40 ‘Draft Regulation’ as a first, essential step in a potentially lengthy and complex legislative process. This was preceded by the establishment of a High-Level Expert Group on AI which produced a comprehensive report41 and the publication of the Commission’s ‘White Paper on Artificial Intelligence’42 which sets out the core elements of a legal framework which, based on the above principles, should follow a risk-based and -differentiated approach (marginal no. 95 ff.). However, those two processes were not seamlessly interconnected.

The structure and approach of the European Commission’s Draft Regulation is guided by established measures in the field of product safety.43 It follows – as recommended by the CAHAI study – what it calls a ‘well-defined’, risk-based and regulatory approach.44 The Commission introduces a comprehensive catalogue of definitions, including a definition of AI as ‘software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outcomes such as content, predictions, recommendations, or decisions influencing the environments they interact with’.45 By combining functional and descriptive elements in the definition, the regulatory scope of the Commission’s draft is broad. While not all types of software are covered, some characteristics of most systems are.

The Commission differentiates between four different risks of AI, namely unacceptable, high, limited and minimal risks.46 Thus, certain practices of Artificial Intelligence are explicitly prohibited but, however, different from the policy field-specific Recommendation of UNESCO with its broadly ethically based approach.

In this context, there has been already been criticism that the text of the draft is ambiguous with regard to the definition of high-risk AI systems and therefore needs to be improved.47 Art. 66 ff. envisages the establishment of a “European Artificial Intelligence Board” to advise and assist the Commission.48 The application and implementation of the Regulation is assigned in Art. 59 to national supervisory authorities to be designated or established by the Member States.49 Art. 60 provides for the establishment of a database for stand-alone high-risk AI systems by the Commission,50 which can be assessed as a very reasonable effort.

It would go beyond the scope of this paper to describe in detail how the EU assess the entire regulatory structure of the Commission’s draft. As points of criticism, however, it is worth mentioning here that the Commission’s draft does not contain sufficient complaint and legal protection mechanisms for individuals and communities affected by AI systems, and that the enforcement system exhibits a certain degree of incoherence.51 However, the draft as a whole represents the world’s first attempt to horizontally regulate AI systems and thus fits into the image of the EU as a big player of digital policy that is unsinkable in its critical approach towards the normative challenges of digitisation – which could admittedly entail risks.

43 Ibid., p. 27 ff.
44 Ibid., p. 27 ff.
45 Ibid., p. 27 ff.
46 Ibid., p. 27 ff.
47 Ibid., p. 27 ff.
48 Ibid., p. 27 ff.
49 Ibid., p. 27 ff.
50 Ibid., p. 27 ff.
51 European Commission, Artificial Intelligence Act, p. 72 ff.
II.4. UNESCO Recommendation on the Ethics of AI: Approach and Added Value

For obvious reasons, the UNESCO Recommendation on the Ethics of AI

Moreover, various aspects of the UNESCO Recommendation deserve particular attention, such as the recognition that the use of AI systems also influences the human mind, the ways in which people think, interact and make decisions, and thus the image of humanity. In its global claim, the UNESCO Recommendation also emphasises the special characteristics and challenges of Low- and Middle-Income Countries (LMICs) – among which so-called Least Developed Countries (LDCs), Landlocked Developing Countries (LLDCs) and Small Island Developing States (SIDS) are highlighted – and thus addresses especially the Global South. The cultures, values and knowledge of the indigenous population, whose recognition, protection and promotion are needed for the development of sustainable digital economies, are particularly emphasised and outlined in the areas of political action (cf. e.g. marginal no. 85).

The UNESCO Recommendation thus points to unique contexts in states and societies that are highly relevant for the development and implementation of AI systems. So far, this aspect has been mentioned at best on the fringes of ethical discussions, especially with regard to the international initiatives described above.

This approach of those initiatives is partly understandable as particularly the CAHAI Recommendation and the European Commission’s Draft Regulation address the European context. However, the international effects of those initiatives must be considered. Especially the regulatory approach of the European Union is being observed very closely internationally, with considerable effects regarding the respective policy and governance initiatives in other states; the EU’s General Data Protection Regulation is a prominent example of this.

The UNESCO Recommendation is particularly legitimised by the multi-stakeholder approach used in its preparation: input from representatives of different societal groups was deliberately sought. Furthermore, in this respect, the UNESCO Recommendation generates added value as its development process can serve as an example for other international standard-setting processes.

II.5. Summary

When it comes to AI, many ethical guidelines overlap in the propagation of some conventional principles, but diverge in the next step when it comes to practical implementation. A profound understanding of AI systems and the consequences of their implementation requires a sound knowledge of society and its processes. This knowledge presupposed, the need for context-sensitive, long-term perspectives on the social embedding of AI technologies could be formulated. Many of the impacts caused by these are determined less by the technology itself than by the conditions in which it is used and the broader economic dynamics with which it is associated. This requires governments to take on a broader and more guiding role.

The particular strengths of the UNESCO Recommendation already mentioned under II include the legitimacy-promoting negotiation process with a multi-stakeholder structure and its character as the first global rule set for an ethically secure use of AI. Significant added value, especially in comparison to the other recommendations, studies and legal acts, lies in the concrete references to the eleven policy fields, including education and science, communication, health and the environment. It is also noteworthy that in the Recommendation, based on ethical rules which are largely derived from human rights, introduces concrete mechanisms and mandates for political action, such as the development of an Ethical Impact Assessment for AI systems.
Conditions for the Implementation of the UNESCO Recommendation on the Ethics of AI in Germany
III. Conditions for the Implementation of the UNESCO Recommendation on the Ethics of AI in Germany

III.1. Introduction

In Germany, AI is receiving a lot of attention in both the public and political debate, especially with regard to regulation and promotion of AI. This chapter describes the conditions for the implementation of the AI Recommendation in Germany in the following policy areas:

- Ethical Impact Assessment (Policy Area 1)
- Ethical Governance and Stewardship (Policy Area 2)
- Data Policy (Policy Area 3)
- Development and International Cooperation (Policy Area 4)
- Gender (Policy Area 6)

In describing the status quo in these policy areas, the following questions will be addressed:

To what extent do (political and regulatory) processes, instruments and governance structures exist in Germany and at the EU level that address AI in these policy areas in accordance with the Recommendation (laws, platforms, funding programmes, bodies, testing mechanisms, standards etc.)? What contents of the Recommendation are still in need of regulation or action in Germany? If there is a need for regulation or action, has it already been addressed in other contexts (e.g. by the Enquete Commission on AI of the German Bundestag)?

In each policy area, the description of the status quo is followed by a brief, exemplary comparison with rules or initiatives in the Global South (especially Africa).

III.2. Ethical Impact Assessment (Policy Area 1)

In the following section, the contents of the UNESCO Recommendation on ethical consequences of AI systems and its current implementation will be examined. As reference documents, the proposal for the AI Act of the EU (EU-AI) and the National AI Strategy of the German Federal Government (NKIS) were taken into consideration in particular. In addition, the Ethical Guidelines for Trustworthy AI of the EU Commission’s High-Level Expert Group on Artificial Intelligence (H-EG-KI), the report of the Enquete Commission on Artificial Intelligence of the German Bundestag (EKKI) and the expert opinion of the Federal Government’s Data Ethics Commission (DEK) were examined.

Number 50 of the UNESCO Recommendation: Ethical Impact Assessments

No. 50 of the Recommendation calls on Member States to introduce an “Ethical Impact Assessment”, which aims to identify and assess the benefits, concerns and risks of AI systems, and to propose appropriate measures for risk prevention, mitigation and monitoring, among other safety mechanisms. The EU-AI, on the other hand, focusses in §§ 30 ff. on the classification of an AI system as high-risk and the measures to be taken on which hazards are to be expected. Ethical aspects such as the impact on fundamental rights are also included here. At the national level, the DEK recommends that the state should be involved in the development of ethical standards for AI, which should then serve as a reference point for impact assessment tools.66 The DEK also recommends involving “civil society actors, data protection authorities, consumer protection experts or spokespersons for organisations representing the parties affected”67 in the standardisation process. The EKKI formulates similar claims.

The Federal Government is therefore obliged to create the basis for an Impact Assessment of AI.68 No. 50 of the UNESCO Recommendation has thus been largely taken up and only needs to be implemented in practice across the board.

Number 51 of the UNESCO Recommendation: Due Diligence and Oversight Mechanisms

No. 51 of the Recommendation calls on Member States and private sector companies to develop due diligence and oversight mechanisms to identify the impact of AI systems on the compliance with human rights, the rule of law and inclusive societies, to prevent or mitigate any negative consequences, and to be accountable for how they deal with them.

The EU-AI provides for a tiered risk management system for AI in Art. 9. This includes measures to safeguard fundamental rights. According to Art. 9(6), tests must be “suitable for fulfilling the intended purpose of the AI system”. Similar provisions are found in the NKIS. It calls for an appropriate control structure and verifiability of AI applications.69 The H-EG-KI emphasises that “human values are central to the way in which AI systems are developed, deployed, used and monitored” which should also serve the respect of fundamental rights.70

In contrast to the UNESCO Recommendation, the national AI Strategy and the planned EU-AI Regulation do not provide for mandatory measures to analyse the socio-economic impact of AI on poverty. The H-EG-KI and the EKKI recognise the potential of AI to bridge the gap between rich and poor71 and to promote economic equality.72 However, concrete measures to explore and use this potential are not mentioned.

Entitlement to the issuance of transparency protocols: The H-EG-KI sees transparency as one of the criteria that a trustworthy AI must fulfill. Adequate documentation must allow ex-post controls by authorities.73 According to Art. 23 of the EU-AI, responsible authorities can demand the release of the protocols according to Art. 20 if this is necessary for the compliance of the AI system with the obligations of the Regulation. The EU-AI does not provide for a right to release of transparency protocols for organisations or (private) individuals.

In contrast, the UNESCO Recommendation emphasises the need to ensure access to information, including information of public interest, held by private entities. This is not only to assess the impact of the socio-economic status on AI use, but also to ensure that the gap between people living in wealth and poverty, as well as the digital divide between and within countries, is not widened by the massive use of AI technologies now and in the future. Member States, private companies and civil society should consider the socio-political and psychological impact of AI-based findings on the decision-making autonomy of humans. AI systems identified as posing potential risks to human rights should be widely tested in the context of Ethical Impact Assessment. Where appropriate, assessments should also include tests under real-world conditions.

The National AI Strategy also recognises transparency as an important building block for AI.74 The EKKI report goes one step further: It states that transparency could also be enforced by means of regulatory measures in order to contain risks.75 People should also be able to demand themselves against discrimination by AI. This requires a right to transparency.76 The DEK takes a similar position regarding a right to transparency of the data used, depending on the risk category of the AI system, necessary.77

The sociological and psychological effects of AI-based recommendation systems need to be researched. The EU-AI prohibits in Art. 5 para. 1 lit. a. any AI system “that deploys subliminal techniques beyond a person’s consciousness in order to materially distort a person’s behaviour in a manner that causes or is likely to cause that person or another person physical or psychological harm”.78 However, just what causes psychological harm is not answered in the proposed Regulation. Nor does it contain any obligation to research this field. The H-EG-KI does not deal with the consequences of recommendation algorithms either.

The EKKI describes interdisciplinairy research into the effects of recommendation algorithms as an urgent task.79 The DEK agrees with this assumption. However, it does not limit its assessment to the effects of recommendation algorithms. The impact of any form of AI should be further researched.80

The EU-AI provides in Art. 5 for comprehensive testing obligations over the entire life cycle of AI. This also includes tests before an AI system is introduced into the market. According to Art. 5 para. 6 of the EU-AI, the test procedures “shall be suitable to achieve the intended purpose of the AI system”. This means that testing in a real environment is not mandatory. Moreover, the national AI Strategy recognises the importance of tests, especially tests in regulatory sandboxes (“Reallabore”).81 The German government supports digital test fields.82

63 Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts (COSID), 2021. Cited as EU-AI.
68 EKKI, p. 76. Cited as EKKI.
69 Ibid. Cited as DEK.
70 EKKI, p. 121. Cited as DEK.
71 NKIS, p. 23.
72 HEG-KI, p. 37.
73 Ibid, p. 9.
74 EKKI, p. 72.
75 HEG-KI, p. 8.
77 NKIS, p. 23.
78 EKKI, p. 464.
80 DEK, p. 186.
81 EKKI, p. 461.
82 Ibid, p. 461.
83 NKIS, p. 20.
84 Ibid.
88 DEK, p. 76. Cited as DEK.
89 Ibid. Cited as DEK.
90 EKKI, p. 121. Cited as DEK.
91 NKIS, p. 23.
Article 64, para. 1 of the EU-AI regulates the powers of the market surveillance authorities. They are given unrestricted access to training, validation and test data sets. In the case of specific violations of fundamental rights, Art. 64 para. 3 provides for special measures. The assessment of an AI as a risk system is also the responsibility of the authorities and is carried out according to the guidelines set out in Title III, Chapter 1.98

The National AI Strategy stipulates the promotion and further development of practical approaches to risk management.99 Furthermore, supervision and enforcement of requirements should primarily be assigned to the supervisory authorities as they would already have competence in this field.100 Ethical Impact Assessments should be transparent and, where appropriate, publicly available.

Neither the Recommendation of the HEG-KI nor the EU-AI provide for the publication of the Ethical Impact Assessment. According to the NCIS, one of the requirements for AI is "a level of transparency (...) that adequately reflects the risks".101 However, it does not provide for specific obligations to publish Ethical Impact Assessments.

Conclusion

There is still a lot to catch up on the part of German legislators and other AI actors when it comes to implementing the UNESCO Recommendation’s provisions on Ethical Impact Assessments. In this regard it would be a promising approach to consider the existing ethical recommendations and documents of the OECD, the Council of Europe and the EU jointly and to distil the respective added values. The UNESCO Recommendation offers a global and holistic framework that proposes important measures across policy areas. The forthcoming EU-AI Regulation, on the other hand, has the advantage of being binding for a small group of states. In a next step, the respective added values should therefore be identified in a comparative review, whereby the UNESCO Recommendation, as the thematically broadest, offers the best basis.

While not included in the draft EU-AI Regulation, other institutions besides UNESCO recommend the right to the publication of transparency protocols for users. Research into the sociological and psychological consequences of recommendation algorithms is also not required at European level, but at national level. Furthermore, no obligation to test AI in a real-world environment at European level exists. Germany could lead the way in implementing the UNESCO Recommendation and demand transparency from AI providers. The digital divide within Germany between the formally educated and the formally less educated, between rich and poor, must also be overcome. The development of an Ethical Impact Assessment tool for AI could be headed by the Office of Technology Assessment at the German Bundestag.

A clear legal framework for an ethical assessment of AI systems is still needed, partly because the EU-AI Regulation is yet to be adopted. The future framework must be able to pursue different objectives in a sensitive way, namely both to introduce effective monitoring mechanisms and to (simultaneously) monitor them again. In particular, public authorities should be required to monitor their own AI use: The DEK recommends an obligation for public authorities to monitor the AI they use102 and to appoint a contact person for the use of AI.103 The EKKI also recommends that public authorities monitor the AI they use and that public authorities’ competences and resources are strengthened for this purpose.104 These points are also put forward in the Recommendation, but extended by more concrete requirements.

A brief, comparative look at India: The world’s largest democracy is one of the few developing and emerging countries with its own AI Strategy.105 Among other things, this strategy stipulates that AI must be subjected to an ethical impact assessment. This is intended in particular to prevent or limit discrimination through AI. The introduction of an ethics council for AI is also recommended.106 A less concrete initiative in the sense of an ethical change of perspective is the focus on Buddhist values in the ethical impact assessment of AI.107

94 DEK, p. 159.
95 Ibid., p. 29.
96 EKKI, p. 184.
In particular § 93 MStV, cf. in this respect also Art. 29 of the basis of an appropriate regulatory frame. With regard to the forms of soft governance of only planned in Germany. Thus, no established UNESCO Recommendation, are regard to their use of AI, as suggested in No. 57 The self-assessments by public authorities with the Recommendation, including ethical stand…perspective and incorporate at best incompletely advanced report on the strategy in December in order to “strengthen the development and expansion of the quality infrastructure on the basis of an appropriate regulatory frame - […] for scientific. AI and to “support civil society networking and involve…ment in the development and use of public good-oriented AI”.

With regard to the forms of soft governance of AI as stipulated in No. 56 of the UNESCO Recommendation, standardisation initiatives have already been launched along the so-called AI standardisation roadmap. However, these initiatives are currently shaped by a primarily technical perspective and incorporate at best incompletely the recommendations, including ethical standards and procedural consultation and review obligations with respect to those affected. The Recommendation can therefore serve as an impetus for further practical initiatives in the future, such as experiments with institutionalised users or stakeholder participation in the context of AI governance.

In a comparable way, European standard-setting initiatives are also influenced by an overemphasis on Serbisseeing at accountable and responsible stakeholders that take part in them. Beyond, the draft Euro…ual human rights and the participation models required to assure this compatibility as put forward in No. 61-63 of the UNESCO Recom…ommendation. It seems likely that, due to the incentive regime of the AI Act, standardisation authorities, in particular the European Commit…ee for Electrotechnical Standardisation, will be granted a strong practical role in “on-the-ground” rulemaking within the scope of the AI Act. Standardisation processes tend to be structurally inaccessible to those affected by AI products and services and to NGOs. This entails the risk of overemphasis on industry interests. Nevertheless, the draft for the EU-AI Regulation reflects some of the basic conditions with regard to ensuring AI systems that comply with human rights and participation requirements as stipulated in the UNESCO Recommendation. In particular, the draft provides for a procedure to report serious incidents and malfunctions (Art. 62) and with regard to the practical conditions for implementation, it is still unclear whether roadmap decisions on AI to be established will be sufficiently equipped to fulfill its role effectively and to protect the public and indivi…duals managed and those affected. Besides the EU-AI draft Regulation, the promotion of AI-based innovation on EU level is tackled by other initiatives, in particular by the Data Governance

Conclusion

The currently prevailing approach of stewardship of AI processes in Germany is clearly expandable to ensure more diversity and participation of stakeholders in the conception and development of AI systems. Especially with regard to the EU-AI Regulation in view of the UNESCO Recommendation, a political and legal definition of which models of participation should be established is still largely based on the “soft” perspective, experiments with (new) institutional structures must be undertaken.

This observation coincides in part with the Enquete Commission’s proposal to expand transparency mechanisms between science, business, politics and civil society in the field of AI research. The Enquete Commission suggests to integrate adapted forms of (anonymous) employee participation into AI governance systems in the area of work. Such participation-oriented institutional and procedural models of stakeholder participation should also be tested beyond the area of “work”. This seems particularly advisable in view of the current overemphasis on standardisation procedures in regulatory approaches (see above). Of course, they may be instructive in view of how to prevent further stakeholder participation in AI governance.

A comparison of the regulatory measures described above with existing regulations in international law reveals that only a limited number of national AI strategies have been intro…duced there. However, some countries have set up task forces and committees that have positioned themselves on the outlined topics, such as Uganda’s “Expert National Task Force on the Fourth Industrial Revolution” (4IR) in whose final report the realisation of the benefits of AI as a tool for progress is valued higher than potential challenges. Also, the question of appropriate (institutional) structures for partici…pation and networking in the field of AI is taken up in the Global South: The final report of the Ugandan 4IR refers, for example, to initiatives for “Centres of Excellence” of the National Association of Software Companies in India, which bring together stakeholders for national development or use certain AI applications together. Uganda wants to transfer this model to the African sector in order to achieve improvements there through a public-private partnership.

34

35
In November 2019, an interim report on the strategy’s implementation was published.\textsuperscript{126} In the area of “data policy”, the report initiated the data infrastructure “GAIA-X”, regular exchanges on AI data protection issues with data protection supervisory authorities on an open data platform.\textsuperscript{127} In addition, the Data Ethics Commission was established in the meantime and its first report was presented in October 2019.\textsuperscript{128} This report emphasizes respect for the rights of individuals involved, use and sharing of data, data quality, information security and transparency.\textsuperscript{129} It also describes in detail, among other things, data rights and corresponding data obligations,\textsuperscript{130} requirements for the use of personal data\textsuperscript{131} and access to personal data.\textsuperscript{132} In December 2020, the update of the AI Strategy was published, which formulates the further steps for the implementation of the strategy.\textsuperscript{133} Also the final report of the Enquete Commission on Artificial Intelligence from October 2020 deals in detail with AI and data.\textsuperscript{134}

The strategy explicitly lists “data use, data security, law and ethics” as one of the fields of action. Its overall intention is to provide an “overview of goals, foundations and fields of action for a national strategy on Artificial Intelligence”.\textsuperscript{135} The KI-Bundesverband is another important institution. In particular, its “data policy” in the field of application of AI in position papers and press releases.\textsuperscript{136}

With regard to funding programmes, the establishment of AI service centers should be highlighted.\textsuperscript{137} Those centers should provide access to data sets and at the same time guarantee data sovereignty. A complete list of funding programmes can be found at “www.ki-strategie-deutschland.de”\textsuperscript{138}. At the European level, a European Commission agenda was published in April 2018, which emphasizes, among other things, access to data and an appropriate ethical and legal framework.\textsuperscript{139} In October 2020, the “White Paper on Artificial Intelligence – A European Approach to Excellence and Trust”\textsuperscript{140} was presented. It also adds it delivered its report on the additional personal data protection and security.\textsuperscript{141} In addition, the draft EU-AL Regulation addresses data and data governance in Art. 10 and describes the quality management system in Art. 17.\textsuperscript{142} Data protection is a central cross-cutting theme of the draft.\textsuperscript{143} In addition, the European Commission’s High-Level Expert Group on AI also published recommendations on policies regarding data economy, data access and protection of affected individuals in a paper in June 2019.\textsuperscript{144} The ethical baseline developed by this group also contain requirements on technical security, privacy, transparency and fairness.\textsuperscript{145}

In addition, an overview of relevant documents and blog posts by stakeholders can be found on the homepage of the European AI Alliance.\textsuperscript{146}

Conclusion

There is a need for action in Germany in the continuation of data protection regulations with AI strategies. This is particularly the case with regard to No. 74 of the UNESCO Recommendation, which advises that already existing data protection regulations be strengthened in order to protect personal data, including particularly sensitive data. In addition, science and business should exchange with each other and work together more intensively and not predominantly conduct projects in parallel. In this respect, AI companies should, in accordance with Nos. 77 of the Recommendation, be encouraged to share the data they collect in order to promote research and innovation.

The initiative “FAIR Forward – Artificial Intelligence for All with Africa and Asia” addresses framework conditions for AI in the Global South.\textsuperscript{149} This is in line with No. 79 of the UNESCO Recommendation. The African Group on AI also published recommendations on the ethical framework conditions for AI development, with a focus on COVID-19.\textsuperscript{150} In addition, the initiative “Artificial Intelligence for Development in Africa (AI4D)” addresses framework conditions for data use of AI.\textsuperscript{151} The focus of these initiatives is primarily on ensuring that the Global South is not left behind on AI issues. In addition, AI should be developed by the countries themselves and not only by the countries of OECD. AI only very few African countries have established AI initiatives.\textsuperscript{152}

The updated version of the AI Strategy also explicitly mentions the report for the creation of framework conditions for AI in the Global South.\textsuperscript{153} This is in line with No. 79 of the UNESCO Recommendation. The African Group on AI calls on Member States to ensure that the use of AI in key sectors of development cooperation is in line with the values and principles set out in this Recommendation. These key sectors include education, science, culture, communication and information, health-care (including food supply, environment and natural resource management and infrastructure, economic planning and growth). The Data Ethics Commission’s report welcomes the choice of a European path that is different from other value systems, cultures and social models.\textsuperscript{154} It is emphasized that the fundamental paradigms of this path should be incorporated even more in global and non-European discourses and contexts. The final report of the Enquete Commission further mentions cooperation between research, business and society as one of the central recommendations for action.\textsuperscript{155}
A working group of the platform “Lernende Systeme” deals with the development field “Health & Care”. In addition, a chapter in the German standardisation roadmap refers to “AI in Medicine.” Besides this, the German AI Association published a position paper entitled “How Artificial Intelligence can contribute to climate protection and sustainability” in February 2021. Among other things, this addresses the creation of a sustainability platform for interdisciplinary cooperation as well as the development fields of energy and agriculture. Furthermore, the funding programme “Digitale GreenTech – Environmental Technology Meets Digitalisation” also focuses on the application fields of resource efficiency, sustainable agriculture and geotechnology.

Germany must, however, increase its commitment to AI research cooperation, especially with LDCs (No. 81 and 82 of the UNESCO Recommendation), in more concrete terms. This commitment should include establishing research and innovation centres and networks with strong participation and leadership of researchers from the Global South. The “White Paper on Artificial Intelligence – A European Approach to Excellence and Trust” calls for cooperation between the public and the private sector as one of the most important measures. The European Commission’s High-Level Expert Group on AI also calls for stronger multi-stakeholder cooperation. Beyond, it also cites the use of ethical guidelines for trustworthy AI in the development cooperation fields of climate protection and sustainable infrastructure, health and education.

An overview of further relevant documents and blog posts by stakeholders can be found on the homepage of the European AI Alliance.

Conclusion
In Germany, there is still a need for action with regard to the policy area “Development and International Cooperation”, especially when it comes to the cooperation of and with private enterprises. In most cases, politics and science are described as cooperation partners, but the private sector is hardly mentioned, especially with regard to development cooperation with the Global South. According to No. 82 of the UNESCO Recommendation, this would be urgently necessary in order to develop AI for specific cultures and contexts, among other things. Cooperation generally focusses on the European continent, which is particularly evident in the mentioning of the “European way”. Yet, individual initiatives and strategies mention the global character of cooperation initiatives, especially with emerging and developing countries. The focus here should also be on the inclusion, in researchers from developing countries themselves, as stipulated in No. 81 of the Recommendation. In addition, there is still a need for regulation or action regarding the overcoming of geotechnological boundaries, as emphasised in No. 83.

The above-mentioned initiative “FAIR Forward – Artificial Intelligence for All with Africa and Asia” is an example of the active inclusion of the Global South. The “Global South – A4COVID” initiative supports interdisciplinary research on COVID-19 in developing countries. In addition, the “Global South – A4COVID” initiative focusses on the Sustainable Development Goals (improving the food chain, education, health and climate change). While in the European context the location of AI businesses is a key topic of the discourse, stakeholders in the Global South increasingly focus on the AI’s potential to foster sustainable development of the society and the environment. Cooperation is usually initiated by organisations and countries that are not part of the Global South.

An important commitment is made at the end of the section in No. 83 of the UNESCO Recommendation. It recalls that all cooperation is based on ethical values and “in full respect of international law”.

III.6. Gender (Policy Area 6)
There are few (political and regulatory) processes, instruments and governance structures in the policy area of gender (and diversity) in Germany and at the EU level that actively address AI. The term “gender” appears exactly once in the German Federal Government’s AI Strategy of November 2018, namely in the area of risk impact assessment of AI in the context of gainful employment. Implicitly, this policy area is negotiated under the premise of non-discrimination. Thus, the general prohibition of discrimination is recalled at various points within the strategy. AI-based decisions, services and products should be made reviewable with regard to potential inadmissible discrimination, according to the strategy. In addition, “transparency, traceability and verifiability of AI systems” should ensure “effective protection against distortions, discrimination, manipulation or other abusive uses, especially in the use of algorithm-based forecasting and decision-making systems”, whereby “discrimination against disadvantaged population groups in developing countries” is also explicitly mentioned. To ensure these imperatives, “the Federal Government is examining the establishment or expansion of government agencies and private audit institutions to monitor algorithmic decisions.” This includes establishing auditing standards and standards for (technological) Impact Assessments. It also involves creating a structure in which all elements of the AI process are disclosed transparently and comprehensively for examination of fundamental rights criteria.

No. 87 of the UNESCO Recommendation also obliges Member States to ensure that the potential of digital technologies and AI contributes to the achievement of gender equality and is fully exploited. Member States must ensure that the human rights and fundamental freedoms of girls and women, as well as their security and integrity, are not violated at any stage of the life cycle of AI systems. Furthermore, the previously described Ethical Impact Assessment should include a transversal gender perspective.

The strategy, measures and current developments are available on the website “www.ai-strategie-deutschland.de”. Also on this online platform, which describes the current political status in Germany, non-discrimination, gender and diversity is not considered a separate field of

168 Ibid.
172 HEQ-Ki, p.32 ff.
175 See Global South A4COVID Program, https://covidsouth.ai/about.
176 See A4Id Africa, https://africa.a4id.ai/about-a4id.
178 Cf. ibid., p. 39.
179 Cf. ibid., p. 38 (translation by the author).
180 Cf. ibid., p. 39 (translation by the author).
181 Cf. ibid. (translation by the author).
182 Cf. ibid., p. 40 (translation by the author).
183 Cf. ibid.
The White Paper mentions the dangers of discrimination by AI with regard to gender already in the first paragraph. This awareness can be observed, though in parts rather superficially, throughout the entire document: “The use of AI can affect the values on which the EU is founded and lead to non-discriminatory, inclusive AI education and training to ensure that AI is non-discriminatory and inclusive”. In the third paragraph, the document is focused on the principles of diversity and non-discrimination as a complex of various, intersectional discriminations with different legal, societal and educational policy connecting factors.

Under the roof of fundamental and human rights, the European AI agenda clarifies the principles of non-discrimination as being part of privacy and consumer protection rights. Besides regulation aimed at transparency, traceability and accountability, there is also a call-for more research to uncover actual patterns of discrimination to prove them scientifically. The European Commission’s recommendation for Artificial Intelligence – ‘A European Concept for Excellence and Innovation’ from 2020 takes up this issue and develops it further.

It very clearly calls for “AI-supported decisions to be regularly reviewed on their non-discriminatory nature.”

On the European level, a first agenda on AI was published in April 2018. It addresses the issue of discrimination in the context of the development phase of AI systems. However, in this very phase, the document defines the problems and provides a possible way: “The development of AI needs to involve more women and people from different backgrounds, including projects focusing specifically with inclusive AI education and training to ensure that AI is non-discriminatory and inclusive”. In the AI 2030 level, gender issues are thus not only considered in terms of the outcome (discrimination) and a simple solution (equal treatment), but are seen as a complex of intersectional discriminations with different legal, societal and educational policy connecting factors.

In Germany only 16% of AI professionals are female (the global share is 22%), and women in the IT sector earn on average 20% less than their male colleagues. To change this it needs, besides stereotype-sensitive early childhood education, explicit support for women and diverse people to sustainably and successfully enter the predominantly white, male technology sector. Also more extrinsic incentives are needed and should be communicated: It is verified that more diverse teams lead to more diverse and also more successful products. Diversity is a need for more viability of women in AI as role models, for example through the dissemination of success stories.

Furthermore, according to No. 88 of the UNESCO Recommendation, Member States must provide earmarked public funds to finance gender-equitable and non-discriminatory AI research. The promotion that national and EU policies include a gender equality action plan.

In its No. 87–93 on gender policy, the UNESCO Recommendation has many similarities with the standards envisaged by the EU, some of which have already been sharpened nationally for the German strategy. The Recommendation goes much further. It includes the special protection of fundamental and human rights of women and ethnic minority groups under the precept of non-discrimination. It also stipulates the promotion of diversity in the AI (development) sector, especially in view of representation and relevance. Discriminatory biases, inherited through technology, should be strongly mitigated.

In its Recommendation, UNESCO lays a strong focus on representation and yet remains on the surface with its demands and measures for action. The UNESCO Recommendation puts great emphasis on maintaining gender equality in AI development and use; Germany still has a lot of work to do in this regard. There is considerable room for improvement with regard to gender equality and diversity in AI development teams in Germany (No. 19). The promotion of women and increasingly also diverse people in the field of natural sciences and STEM has been legalized in Germany for the federal level for several years and supported by various funding projects in the field of education and vocational training (additionally, cf. HIIG AI Lab, Positionspapier zum Roundtable “KI und Frauen*, December 2020, https://www.plattform-lernende-systeme.de/ #21_1683#2).

In Germany only 16% of AI professionals are female (the global share is 22%), and women in the IT sector earn on average 20% less than their male colleagues. To change this it needs, besides stereotype-sensitive early childhood education, explicit support for women and diverse people to sustainably and successfully enter the predominantly white, male technology sector. Also more extrinsic incentives are needed and should be communicated: It is verified that more diverse teams lead to more diverse and also more successful products. Diversity is a need for more viability of women in AI as role models, for example through the dissemination of success stories.

Furthermore, according to No. 88 of the UNESCO Recommendation, Member States must provide earmarked public funds to finance gender-equitable and non-discriminatory AI research. The promotion that national and EU policies include a gender equality action plan.

In its No. 87–93 on gender policy, the UNESCO Recommendation has many similarities with the standards envisaged by the EU, some of which have already been sharpened nationally for the German strategy. The Recommendation goes much further. It includes the special protection of fundamental and human rights of women and ethnic minority groups under the precept of non-discrimination. It also stipulates the promotion of diversity in the AI (development) sector, especially in view of representation and relevance. Discriminatory biases, inherited through technology, should be strongly mitigated.

In its Recommendation, UNESCO lays a strong focus on representation and yet remains on the surface with its demands and measures for action. The UNESCO Recommendation puts great emphasis on maintaining gender equality in AI development and use; Germany still has a lot of work to do in this regard. There is considerable room for improvement with regard to gender equality and diversity in AI development teams in Germany (No. 19). The promotion of women and increasingly also diverse people in the field of natural sciences and STEM has been legalized in Germany for the federal level for several years and supported by various funding projects in the field of education and vocational training (additionally, cf. HIIG AI Lab, Positionspapier zum Roundtable “KI und Frauen*, December 2020, https://www.plattform-lernende-systeme.de/ #21_1683#2).

In Germany only 16% of AI professionals are female (the global share is 22%), and women in the IT sector earn on average 20% less than their male colleagues. To change this it needs, besides stereotype-sensitive early childhood education, explicit support for women and diverse people to sustainably and successfully enter the predominantly white, male technology sector. Also more extrinsic incentives are needed and should be communicated: It is verified that more diverse teams lead to more diverse and also more successful products. Diversity is a need for more viability of women in AI as role models, for example through the dissemination of success stories. 

In Germany only 16% of AI professionals are female (the global share is 22%), and women in the IT sector earn on average 20% less than their male colleagues. To change this it needs, besides stereotype-sensitive early childhood education, explicit support for women and diverse people to sustainably and successfully enter the predominantly white, male technology sector. Also more extrinsic incentives are needed and should be communicated: It is verified that more diverse teams lead to more diverse and also more successful products. Diversity is a need for more viability of women in AI as role models, for example through the dissemination of success stories. 

In Germany only 16% of AI professionals are female (the global share is 22%), and women in the IT sector earn on average 20% less than their male colleagues. To change this it needs, besides stereotype-sensitive early childhood education, explicit support for women and diverse people to sustainably and successfully enter the predominantly white, male technology sector. Also more extrinsic incentives are needed and should be communicated: It is verified that more diverse teams lead to more diverse and also more successful products. Diversity is a need for more viability of women in AI as role models, for example through the dissemination of success stories.

In Germany only 16% of AI professionals are female (the global share is 22%), and women in the IT sector earn on average 20% less than their male colleagues. To change this it needs, besides stereotype-sensitive early childhood education, explicit support for women and diverse people to sustainably and successfully enter the predominantly white, male technology sector. Also more extrinsic incentives are needed and should be communicated: It is verified that more diverse teams lead to more diverse and also more successful products. Diversity is a need for more viability of women in AI as role models, for example through the dissemination of success stories.

In Germany only 16% of AI professionals are female (the global share is 22%), and women in the IT sector earn on average 20% less than their male colleagues. To change this it needs, besides stereotype-sensitive early childhood education, explicit support for women and diverse people to sustainably and successfully enter the predominantly white, male technology sector. Also more extrinsic incentives are needed and should be communicated: It is verified that more diverse teams lead to more diverse and also more successful products. Diversity is a need for more viability of women in AI as role models, for example through the dissemination of success stories.

In Germany only 16% of AI professionals are female (the global share is 22%), and women in the IT sector earn on average 20% less than their male colleagues. To change this it needs, besides stereotype-sensitive early childhood education, explicit support for women and diverse people to sustainably and successfully enter the predominantly white, male technology sector. Also more extrinsic incentives are needed and should be communicated: It is verified that more diverse teams lead to more diverse and also more successful products. Diversity is a need for more viability of women in AI as role models, for example through the dissemination of success stories.
Alongside accessible education for all ages, there is a need for a politically driven, serious structural change in the (business) logic of IT and AI companies. The current imbalance in representation will not be fixed through isolation alone; strong incentives for structural and systemic change within the organisational cultures of companies are needed. Furthermore, a re-thinking of AI development in an intersectional perspective is needed to make different patterns of exclusion visible and to overcome them.

Conclusion

Member States should ensure that AI does not exacerbate the large gender gaps that already exist in the analogue world in various fields, but rather eliminate these differences. These gaps include: the gender pay gap; unequal representation in certain professions and activities; lack of representation in executive positions, supervisory boards or research teams; the educational gap; the gap in access to digital content; the unequal distribution of unpaid work and care responsibilities. Member States should also, according to No. 90 of the UNESCO Recommendation, ensure that gender stereotypes and discriminatory biases are not incorporated into AI systems, but that they are identified and proactively eliminated. As one group of experts recently formulated: “AI development must be thought of intersectionally in order to break through manifest systems of exclusion and make rigid category systems visible. This requires not least an intensive examination of the question of access to material and immaterial resources, opportunities for access and participation, opportunities for education and involvement.”

According to the OECD, there are initiatives on AI and gender in six African countries. Of course, there are individual initiatives, partly also funded by Germany, that push the issue, such as “FAIR Forward – Artificial Intelligence for All with Africa and Asia”. However, these initiatives only deal marginally with the topic of gender, diversity and non-discrimination. In the initiative “Artificial Intelligence for Development in Africa” (A4D Africa), the promotion of inclusive, diversity-sensitive AI is granted a stronger focus, but has not yet been translated into a regulatory framework.

With regard to the commitment that gender equality must not be violated at any stage of the life cycle of AI systems, Germany is also still at the very beginning (No. 90 of the Recommendation): Ensuring that gender stereotypes and discriminatory biases are not incorporated into AI systems but identified and proactively eliminated, will lead to substantial changes in the state’s approach to AI-based decision-making processes.

Equal treatment also relates to data policy, as many data with a gender dimension or those related to sexual orientation are very sensitive. In No. 74 of the Recommendation, Member States commit to providing special protection for sensitive data. These include: Data on offences, criminal proceedings, convictions and related security measures; biometric, genetic and health data; personal data such as data on race, colour, descent, sex, age, language, religion, political opinion, national origin, ethnic origin, social origin, economic or social circumstances of birth or disability and other characteristics.

Concerning “remote biometric identification” it becomes at least implicitly apparent how highly sensitive these technologies are in relation to other fundamental rights. Clearly, any infringement of these fundamental rights must be urgently prevented. Even more concrete is the European Commission’s proposal for an AI Regulation. Here, Articles 13, 15, 17, 33, 35–39, 44, 45 and 47 take up in great detail what is already laid down in the White Paper: namely, the unconditional prevention of the reinforcement of (implicit) biases by AI used in highly sensitive areas and trained with datasets that are not free of intersectional discrimination.

204 Ibid., p. 11.
205 Ibid., p. 12.
206 Ibid. (translation by the author).
210 Ibid., p. 25.

Conclusion
IV. Conclusion

Most ethical guidelines on AI propagate general principles and do not aim at practical implementation. Yet, the UNESCO Recommendation on the Ethics of AI is the first international legal text negotiated in an open and inclusive process that translates ethical principles into concrete tasks for policy-making. The Recommendation includes globally accepted ethical standards for AI technologies with full respect for human rights.

It is remarkable that all 193 UNESCO Member States have committed to the Recommendation. The preceding two-year development process, in which non-state actors were also able to contribute their voices, added significant value in terms of legitimacy. As regards content, the holistic approach of the Recommendation stands out. On a general level, it becomes clear that the Recommendation considers AI ethics a task for society. In order to fulfil it, citizens must be made aware of the potential and risks of AI. On the policy level, the Recommendation also defines AI ethics as a directly applicable set of normative standards. Thus, it formulates concrete tasks for national action on AI development and use in eleven policy areas, including education and science, communication, health and the environment. The focus is not on abstract demands, but on concrete improvements.

The Recommendation makes use of the advantages of ethics over law, particularly with regard to policy development and implementation, and in clarifying legal norms, such as equal treatment obligations, which are extended to AI applications. The reference to ethics as well as the soft law status of a Recommendation allowed both a strong human rights language and a focus on practical implementation to be embedded in the document. This would not have been possible with the objective of a “hard” global AI law, as normative orders are often only just being created. The consideration of previous “blind spots” in AI law, such as the impact of AI on gender, sustainability, development policy or education is also noteworthy.

Furthermore, the Recommendation defines procedural and transparency obligations: Specifically, it calls for Ethical Impact Assessments for the development and use of AI systems, which in turn should occur in a multi-stakeholder-based, open and responsive process.

A further added value of the UNESCO Recommendation is its reference to cultural aspects possibly affected by the use of AI systems. These include potentially positive effects such as the preservation of endangered languages, but also negative effects like new access barriers on cultural digital platforms. For the first time, the Recommendation also focusses on the particular dynamics that countries of the Global South may face in the use and development of AI systems. It not only points out the potentially higher vulnerability of developing countries, but also highlights their ability to contribute to fairer AI use and better AI regulation. By emphasising the importance of certain marginalised groups and the need for their protection, the UNESCO Recommendation can provide an important impetus for the EU’s legislative process. A sufficient level of protection can thus be ensured through suitable regulatory instruments and mechanisms.

The UNESCO Recommendation on the Ethics of AI is the first global consensus on the ethical development and use of Artificial Intelligence. It opens a unique opportunity to guide AI development in an ethical manner and thus to actively shape it in the sense of human (rights). At this point, it is now up to the Member States to fulfil their obligation under to international law and implement the Recommendation. The laws, programmes and funding instruments that are now adopted on the basis of the Recommendation will, in retrospect, be the yardstick for its success.