THERESA ZÜGER, LENA WINTER, LAURA STATE

Impact Matters!

Overview of the development and application process of the Impact AI Method: a holistic impact assessment for running AI projects

ABSTRACT

The overall rapid increase in the use of artificial intelligence technologies (AIT) is linked to various initiatives that propose AI 'for good'. However, a holistic, in-depth understanding of these AI use cases and their implications for people, society, and the planet remains lacking, beyond the respective projects' self-descriptions. To close this gap, we present a framework for a holistic yet manageable approach for running AI projects to assess their impact. We call our approach the Impact AI Method, as we not only want to assess the *impacts of AI* use cases but also want to *impact AI and its use* in practice and in public debate. The Impact AI Method evaluates concrete AI use cases in terms of their contribution to sustainability and public interest, reflecting these two concepts as well as established international guiding principles such as democratic values and human rights. The interview- and survey-based methodological approach combines and extends existing assessment approaches, incorporating a catalogue of 25 assessment criteria to evaluate the results. The Impact AI Method is intended as a reusable blueprint to inform public

debate about AI 'for good' claims and to support the creation of evidence and transparency for AI systems that purport to contribute to just and sustainable development.

KEYWORDS

Artificial intelligence, impact assessment, sustainability, public interest, fundamental rights

CITATION

Züger, T., Winter, L. & State L. (2026). Impact matters! Overview of the development and application process of the Impact AI Method: a holistic impact assessment for running AI projects. HIIG Discussion Paper Series 2026-10. 22 pages. https://doi.org/10.5281/zenodo.23887282

LICENCE

This work is distributed under the terms of the Creative Commons Attribution 4.0 License (International), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (https://creativecommons.org/licenses/by/4.0/). Copyright remains with the authors.

FUNDING

The Impact AI project is funded as part of the Change! Fellowship Program of the Volkswagen Foundation.

CONTENTS

1 INTRODUCING THE IMPACT AI METHOD 3
2 RELATED WORK 4
3 NORMATIVE AND THEORETICAL BACKGROUND 5
4 METHOD DEVELOPMENT 7
5 METHODOLOGICAL APPROACH AND OVERVIEW OF THE IMPACT AI METHOD MODULES 10
5.1 The preparation phase 11
5.2 The technical survey 12
5.3 The technical follow-up interview 12
5.4 The organizational interviews 12
5.4.1 AI project impact and governance interview 13
5.4.2 AI project sustainability interview 14
5.4.3 AI project risk and mitigation interview 15
5.5 External impact interviews 15
5.6 Data analysis 16
5.7 Normative assessment 16
6 DISCUSSION 19
7 CONCLUSION AND OUTLOOK 22
8 REFERENCES 22
9 APPENDIX 26

1 INTRODUCING THE IMPACT AI METHOD

This discussion paper introduces a structured and replicable approach for an impact assessment of deployed AI use cases, aiming for a holistic assessment. The motivation for developing the Impact AI Method stems from a research gap we identified in previous work (Züger & Asghari, 2023; Züger et al., 2022). In the past decade, the number of AI applications aiming to use AI for 'good' seems to be rising, and with them the claim about a positive impact and potential of AI to serve society (Tomašev et al., 2020; Floridi et al., 2020; Cowls et al., 2021; Chui et al., 2018; European Union, 2025). At the same time, a deeper understanding of these projects, their design choices, and their intended and unintended impacts on society and the planet is still lacking. The status quo is far from a comprehensive understanding that examines the effects of AI use cases intended to do good beyond their public self-descriptions. The approach we are proposing with the Impact AI Method focuses on the impacts of individual AI use cases in detail, but with the intention to support the widespread application of the method and a growing number of assessed use cases. With the in-depth assessment of a multitude of individual cases, the field will gain a more nuanced and accurate understanding of the reality of AI for good applications.

While existing works on impact assessment of AI are helpful and relevant, our approach takes on the task of combining, extending and aligning these into a holistic and replicable approach, providing a transparent overview and evaluation of the impacts of specific AI for good application. Overall, the Impact AI Method is designed to answer the research question, *which implications does a given AI use case (X) have for people, society, and the planet?* To answer this question, the Impact AI Method **provides a framework for a holistic yet manageable approach for running AI projects to assess their impacts**. It allows us to answer, in a more general way, which positive outcomes and potentially which harms and negative impacts specific use cases of AI may have. By manageable, we mean that the method requires the framework to be comprehensible, applicable by practice-oriented evaluators, and demands a time commitment that is realistic for organizations running such projects (often NGOs). The answer to this question is not given arbitrarily, but in reference to existing norms and values, such as the Universal Declaration of Human Rights, the United Nation's understanding of democracy, a multidimensional account of sustainability, as well as consideration of what it means for a technology to serve public interests, as for instance articulated for the AI Act of the European Union.

Nevertheless, answering this question requires breaking this overarching question down into numerous smaller aspects as well as an awareness of the existing limitations of providing a conclusive or final answer. With this awareness, we see the respective results of our impact assessment as a transparent interpretation of the cases we researched. We are convinced of its value for research and practice as, on one hand, we gained insights that have the potential to enhance the evidence and discussion in academic debate, and, on the other hand, received positive feedback from the cases themselves that the offered reflection process is helpful for internal awareness and improvements. We call our approach the Impact AI Method, as we not only want to assess the *impacts of AI* use cases but also want to *impact AI and its use* in practice and in public debate.

The methodological approach presented here draws on a number of existing impact assessments and takes inspiration from existing methods for documenting AI systems, AI auditing frameworks, and a corpus of research aimed at determining and evaluating the societal and environmental effects of AI systems. Further, we draw on a combination of the concepts of public interest and sustainability, as we see these as a theoretical framework with which various aspects of impact can be comprehensively operationalized. Alongside this foundational paper laying out the related work and theoretical background, method development process, methodological approach and modules as well as their application in practice, we publish materials to apply the method, including questionnaires and templates (the Impact AI Method and its materials are available here).

2 RELATED WORK

Previous research exists that attempts to analyse AI-supported projects for a positive outcome, e.g. in the support of SDGs (Nasir et al., 2023; Vinuesa et al., 2020; Chui et al., 2018; Saetra, 2021; Palomares et al., 2021; Cowls et al., 2021; Selgas-Cors, 2025; Hussein & Sid Ahmed, 2025), often posing this question on a meta level and therefore utilizing scientific repositories and research papers on related AI use cases or research projects. Nasir et al. (2023), for instance, have provided overviews of how projects align with the SDGs in their descriptions, and how they appear to support or hinder them at this descriptive level. However, this perspective does not necessarily capture the actual impacts of individual projects. Also, this meta-level perspective leads to an ambiguous result, showing that while some SDGs seem to be furthered, others are hindered by the use of AI in specific contexts.

Beyond the research on this meta perspective, impact assessments of concrete use cases have a long history as a reflective and practice-oriented approach and assessed impacts in regard to social and environmental dimensions, human rights as well as privacy, data protection, or ethics long before AI emerged. The International Association for Impact Assessment defines impact assessments as “a structured process for considering the implications for people and their environment of proposed actions, while there is still an opportunity to modify - or even, if appropriate, abandon - the proposals” (International Association for Impact Assessment, n. d.). Today, also for AI, numerous impact assessments exist (Stahl et al., 2023), while there is no agreement yet on a necessary standard in content, structure, and implementation. In their systematic review, Stahl et al. (2023) identified 38 different AI impact assessments, each targeting different goals and outlining different approaches. Some of these assessments are products of academic work; others are brought forward by public sector bodies, standardisation bodies, civil society organisations or even corporate organisations. Also, their motivations for the respective impact assessments differ: For some, the motivations are safeguarding the benefits of AI (High-Level Expert Group on Artificial Intelligence [AI HLEG], 2020), assessing system acceptability (Reisman et al., 2018), and promoting trustworthy AI (Calvo et al., 2020; AI HLEG, 2020). Our motivation is most similar to that of IEEE (2020) and Raji et al. (2020). Here, the understanding of the impacts of AI is emphasized, for instance, in particular related to impacts on human well-being throughout the life cycle of AI systems and identifying intended and unintended uses (IEEE 2020) or anticipating negative impacts and how to mitigate them (Raji et al. 2020). These assessments are usually designed for AI use cases across domains and regardless of their explicit goals, which may be the automation of a process or driving efficiency.
Our approach starts from a slightly different angle. Our impact assessment targets a reflection on the

realization of the claimed impacts for 'good' alongside a broader normative framework, as we are particularly interested in better understanding if the use of AI can uphold the promises and claims that are brought in connection with it. Similar to other impact assessments, we aim to achieve knowledge that helps to better AI use cases: specifically, the ones we research, but also provide transferable knowledge to other use cases by pointing to weaknesses, strengthening accountability, creating transparency, and providing an assessment framework that raises the most relevant questions for improvement. While some impact assessments target the critical cases, for which negative impacts, such as potential harm, are to be expected (Stahl et al., 2023), our approach is specifically tailored to assess projects with regard to the impact claims they make themselves, and assess the evidence for these impacts in reflection of established international values.

Stahl et al. (2023) grouped the corpus of reviewed impact assessments into six categories: 1) human rights, 2) ethics, 3) data protection and privacy, 4) security, 5) safety and 6) environmental impacts. The authors note that human (or fundamental) rights are the issue most frequently referenced. Many existing works on the impacts of AI are rather narrow and focus, for instance, on environmental sustainability only (Luccioni et al., 2024), or on privacy-related issues specifically (Clarke, 2009). Each issue that is focused on, whether it be ethical, sustainable, or privacy-centered, determines which aspects of a given use case are considered in regard to impacts, and which are left out or paid less attention to.

Focused on the general research question *which implications does a given AI use case (X) have for people, society, and the planet?* Our position is that to assess this research question, one needs to cover not only the above-mentioned six issues, but even more. Our approach adds a more political lens that allows us to ask about impacts on society as a whole, on democratic values, such as public interests, and on sustainability in all its dimensions: social, environmental, and economic. With this, our method has a two-eyed focus: it assesses the AI use case itself (meaning the system, data, model, and architecture and its design) as well as the societal purpose (and its achievement) which the AI system is intended to support. Only by allowing a gaze for this "broader picture" can one really answer the question of AI societal impacts holistically. Certainly, an assessment of these political dimensions cannot be achieved conclusively, but we believe that for a more evidence-based and nuanced debate on the concrete impacts of AI use cases, we need to start asking about impacts on political dimensions of AI and articulate our findings as much as possible to enhance public discourse.

3 NORMATIVE AND THEORETICAL BACKGROUND

To answer our research question, we also take a normative approach, which requires going beyond a descriptive outcome. To operationalize the impacts for people, society, and the planet, we grounded our method in established normative ideals that lead to a substantial answer to our research question, allowing us to evaluate the results in relation to existing articulations of these norms. Given the fact that our approach is applicable to AI projects in diverse international contexts, we decided to ground our work in the values and principles of international organizations, primarily the UN. The value basis of our framework is the Universal Declaration of Human Rights (United Nations [UN], 1948), the UN's definitional work on the concept of Democracy (UN, 2009), as well as the articulation of sustainability by the World Commission on Environment and Development (1987), which was later established as the basis

of the UN's Sustainable Development Goals.
Sustainability is an intentionally recognized guiding concept which articulates negotiated ideals and goals around intergenerational and intragenerational justice for people and the planet (World Commission on Environment and Development [WCED], 1987; UN General Assembly, 2015). Basing our method on conceptual works on sustainability allows us to focus on impacts on people and the planet in a multitude of aspects, integrating the established dimensions of social, environmental, and economic sustainability (UN General Assembly, 2015).
Within the UN's work, the concept of the public interest can be seen as an underlying implicit goal of the SDGs and democracies as well (Gale, 2022; Chelimsky, 2014; Lewis, 2006). Public interests are for instance articulated alongside the AI Act, stating that "[the AI Act as] a Union legal framework [is] laying down harmonised rules on AI [...] needed to foster the development, use and uptake of AI in the internal market that at the same time meets a high level of protection of public interests, such as health and safety and the protection of fundamental rights, including democracy, the rule of law and environmental protection as recognised and protected by Union law" (Regulation (EU) 2024/1689 (Artificial Intelligence Act): 2). Articulating these values and principles as a normative foundation allows us to bring the collective long-term survival and well-being of a public and the planet as well as the procedural qualities of democratic negotiations of these interests into focus (Bozeman, 2007). It thereby widens the scope of our research to pay attention to the quality of participatory and procedural aspects and allows a political assessment of AI use cases, which we deem necessary.

In agreement with Weidner (2002), we see the combination of sustainability and public interest(s) as necessary for future decision-making, which enables sustainable and just development. Weidner describes a productive tension between the two concepts, which means that they enrich each other both theoretically and in practice. We agree with him that both concepts are to be evaluated on an equal and not hierarchical level. Each concept on its own appears "unsuitable in political and social practice for solving pressing global problems and shaping a global civil society" (Weidner, 2002). Therefore, as Weidner argues, a "regulative double-idea" is essential: the combined consideration of public interest and sustainability is "unavoidable, since under current global development trends there can be no public interest without sustainability, but also no sustainability without public interest" (Weidner, 2002). At the same time, in the spirit of complementarity, it must incorporate the seemingly irresolvable tension between the two concepts and guide political learning and decision-making. On the one hand, sustainability and public interest have a strong overlap; for example, both aim for collective well-being and equity and are based on shared values such as justice, human rights, and democracy. However, the two concepts also differ: while public interest tends to provide arguments for a specific, rather local community and interests that have meaning at a given time–even though a long-term perspective might be relevant–sustainability inherently builds on a perspective that aims for justice between current and future generations and tends towards a global viewpoint. Furthermore, sustainability as a multidimensional concept is already well spelled out in societal practices but has normative weaknesses in its internal trade-offs (Kroll et al., 2019), for example, when prosperity is served at the expense of social or environmental sustainability. Public interest conceptions are a counterpoint as they require the negotiation of all relevant societal interests equally. Additionally, the public interest perspective puts focus on the question of purposeful and meaningful action and introduces emphasis on the procedural quality of how this purpose is achieved by

requiring a participatory and deliberative approach to the negotiation of interests. A good example of what we mean by productive tension is the question of human-centricity, which is an issue for debate in sustainable politics overall, but also AI specifically.

Theories of public interest commonly center around human survival and well-being (today and in the future). Environmental issues can be captured as public interests and even negotiated in public interest litigation (Hui & Karim, 2025). However, public interest negotiation addresses these environmental concerns by articulating interests for a human community. The interests of animals and nature are only captured indirectly.

In contrast, multidimensional understandings of sustainability take into account the environmental well-being of life on land and life in water on their own behalf, seeing "intrinsic value in biodiversity"[1], without it serving humanity. This tension between a human-centric and a more holistic view on livelihood and well-being of our planet also shows in regard to AI systems specifically. While on many occasions human-centric design of AI systems is seen as a normative quality (Kazim & Koshiyama, 2020), this perspective is not without critique (Ryan, 2025) or alternative views (Flaksberg, 2024; Munn, 2024). Critics deem the human-centric approach to technology to be "too simplistic, because it does not account for the complex entanglements of Earth's ecosystems" (Foth et al., 2021).

This example of a tension between the two perspectives shows how they surface the need for a necessary process of global and local negotiation to find practical solutions to partly conflicting objectives and interests. We suggest that both perspectives are highly meaningful to a holistic understanding of the impact of concrete AI systems in practice and therefore use public interest and sustainability as a theoretical framework for our methodological approach. Building on this theoretical foundation, in the following we will lay out the process of method development and explain which existing approaches are influential and integrated into the overall framework.

4 METHOD DEVELOPMENT

Our method development started with iterative cycles of desk research, screening, and adopting existing methods. We collected and reviewed relevant literature, focusing on impact assessment approaches and documentation of AI systems, relevant international AI frameworks and standards, as well as pertinent literature on sustainability, public interest and AI.

From the reviewed literature, we extracted a first set of potential indicators of the Impact AI Method. After arranging the indicators and questions into separate interview guides or the technical survey, we closed critical gaps, removed redundancies, and addressed understandability issues.

Inspirational work for the development of the qualitative interview guidelines for our method stems from a multitude of sources, as many existing approaches were relevant to our goal of creating a holistic approach. One relevant strain of work is **ethical guidelines or ethical considerations on AI** (Jobin et al., 2019; Cowls et al., 2021; Floridi et al., 2020). Ethical principles for AI for good have been propagated and discussed for more than the last decade and are surely relevant. Many aspects of these are integrated into our assessment method as a normative foundation and in qualitative questions. Another source of inspiration was therefore more political, theoretical, and empirical works on the question of how AI could

[1] https://sdgs.un.org/topics/biodiversity-and-ecosystems

serve the public (Züger & Asghari, 2023; Züger et al., 2022).

We understand **guidelines by international institutions** on the responsible, ethical, or trustworthy use of AI as an operationalization of not only the social sustainability of AI but also public interest-oriented values for AI, such as, for instance, the Principles for Trustworthy AI by the Organisation for Economic Co-operation and Development [OECD] (2024), the Ethical Impact Assessment (EIA) by the United Nations Educational, Scientific and Cultural Organization [UNESCO] (2023) or the EU Assessment List for Trustworthy AI (ALTAI) (European Commission, 2020). While our method overlaps and is synchronized in parts with the guidelines and approaches mentioned, it also differs in scope, as, for instance, trustworthy AI does not address the question of realized intended outcomes of AI use cases. Also, not all of these provide concrete questions for an assessment, and the EIA in particular is designed as a self-assessment to be conducted before the AI system is running. We instead are interested in impacts of running systems and assessments conducted by or at least with external third parties.

Another important strand of literature is **international standards**, such as the ISO/IEC standard 42005 on AI System Impact Assessments (ISO/IEC 42005:2025), ISO 26000 on guidance on Social Responsibility (Deutsches Institut für Normung, 2021), or the Global Reporting Initiative (GRI) standards on Sustainability Reporting (2024). The latter two do not have a specific AI focus but include important topics such as human rights, protection of the environment, or community engagement. However, they represent recommendations for action and not formulated assessment questions. The scope of ISO/IEC 42005:2025 includes, among others, the purpose of the AI system, intended as well as unintended use and misuse of the system, and impacts on affected parties. It also covers technical information on the AI system, data, or algorithm. These topics overlap with other considered guidelines or assessments and are therefore all important and included in our approach.

Several case studies on assessing the sustainability of AI exist and were informative to our methodological approach (Strubell et al., 2019; Luccioni et al., 2024; Luccioni & Hernandez-Garcia, 2023). They mainly focus on the environmental dimension of sustainability and are less encompassing with regard to societal and economic aspects of sustainable application of AI. The most comprehensive approach to assess the sustainability of AI is by Rohde et al. (2024), putting forward 19 assessment criteria for AI systems, covering the three sustainability dimensions (this and the aforementioned work were also particularly relevant for the development of the technical survey). We integrated the approach by Rohde et al. (2024) into our method. Many assessment questions were not always as in-depth or open enough to include contextual details, so we adjusted them for our method. Furthermore, the project focused on supervised models, while our approach can be applied to any type of ML model. In addition to these AI sustainability criteria and indicators, the most relevant sources inspiring the interview questions were the EIA (UNESCO, 2023) together with the AI Supply Chain Impact Framework (AI + Planetary Justice Alliance, 2025) and the Fundamental Rights Impact Assessment (FRIA) template developed by the European Center for Not-for-Profit Law and the Danish Institute for Human Rights (2025). The interview questions also operationalize existing work on Public Interest AI principles (Züger & Asghari, 2023; Züger et al., 2022).

As with other methods, one important step of any assessment is to gain an accurate understanding of the technology in place by requesting documentation (Stahl et al., 2023). In several iterations, we developed three comprehensive survey templates, called the AIMS templates, standing for “Artificial Intelligence through the lens of Multidimensional Sustainability templates” (State et al., 2026). The documentation

stemming from the responses to these templates allows one to understand the AI system, model, and used data, and inherent design decisions. The AIMS templates were inspired by work on AI documentation such as the model cards framework (Mitchell et al., 2019), the datasheet framework (Gebru et al., 2021), as well as the TechOps (Lucaj et al., 2025) framework. The latter are documentation templates for compliance with the European AI Act. To develop three separate surveys–which together form the AIMS templates–was inspired by Lucaj et al. (2025). The criteria and indicators operationalizing multidimensional sustainability developed by Rohde et al. (2024) are reflected in the template. In focus here are the technical aspects and details of the AI system. Additionally, there is also a growing body of work on how energy use and other aspects of environmental sustainability can be documented for an AI model (Fischer et al., 2022; Henderson et al., 2020), which was relevant for the survey questions. The AIMS templates at this moment are more comprehensive than any of the approaches we have seen: none of the previously existing templates for system documentation cover the full AI system, i.e. the model and the dataset, but also hardware structure and the embedding of the AI system. Also, even if the templates cover more than only parts of an AI system, they focus only on specific types of impact or on legal compliance.

Our effort to understand projects relative to public interest(s) led us to a stronger focus on the roles of different stakeholders, questions around the participation and involvement of affected groups, and how design decisions reflect public interest principles. Also, it led us to focus on evaluating both evidence and self-reports on impact for the individual AI project's self-defined societal goal, which, to our knowledge, is not included in other methods at this point.
Integrating a multidimensional understanding of sustainability inspired how the indicators and questions of the interviews and the technical survey are clustered and resulted in them being grouped along the three dimensions of sustainability.

In parallel, we developed the catalogue of the 25 assessment criteria and associated questions. The assessment criteria are aligned with the "Principles for the Ethical Use of Artificial Intelligence in the United Nations System" (UN, 2022), which is a list of principles (in line with many other standards) and also explanations of what their realization would mean in practice for the UN. We opted for a close alignment and operationalization of these UN principles for the UN's own use of AI, as this framework is largely aligned with many other existing value-based frameworks and has an international foundation. Also, it is already practice-oriented and includes thoughts on how these principles should be spelled out in practice. We extended the catalogue with impact-oriented assessment criteria to more deeply reflect aspects of public interest and a multidimensional understanding of sustainability in combination (see Table 3 below).

Throughout the development process, there were several (intentionally created) channels from which we received feedback that altered and improved the method. Before conducting the first impact assessment, we applied the method to a test case which was an AI project based at our institution. This helped improve the method regarding interview flow and the understandability of questions. Additionally, we received important feedback from our research council[2], which had significant influence on the further development leading to the second version of the method. The research council consists of eleven members from

[2] https://www.hiig.de/en/impact-ai-project-council/

academia, civil society, and industry and covers different expertise, such as sustainability and AI, human rights, technology and law, and transformation research. In particular, based on the council's feedback, we have aligned our method more closely with other existing impact assessment methods. Firstly, we have incorporated a template for a Fundamental Rights Impact Assessment (FRIA) developed by the European Center for Not-for-Profit Law and the Danish Institute for Human Rights (2025). We have adopted a FRIA triage–a shorter version based on this template–which helps to operationalize the question of potential harmfulness of the AI system and indicates whether a full FRIA may be indicated. Secondly, we have complemented and refined our questions related to data protection to better understand and assess potential issues around the collection and processing of personal data within the AI system. To do so, we drew on existing work and aligned questions with a manual on Data Protection Impact Assessments (DPIA) (Martin et al., 2020); however, we do not provide a full DPIA.

We also learned from the first practical applications of the method to AI use cases. The main iteration from that experience was to design and include the technical survey, based on the AIMS Templates, in the data collection approach of the method. In a first application, the method still contained two semi-structured technical interviews, which were adapted to the technical survey with three sub-surveys in the current second version. This more structured approach to technical documentation of the system has several advantages: first, it gives the project owners of the AI use case more flexibility to choose a particular time to fill out the surveys. Secondly, it allows interviewed parties to look up numbers and answers they do not know by heart about their system, which is harder in a verbal interview setting. Thirdly, the documentation produced is more comprehensible and structured, and finally, it also allows the project owners of a use case to further use this documentation internally, as the survey can be sent back to them as a CSV file.

While the method so far has been applied in four different case studies, we will further iterate the method design. We chose this iterative approach to be able to learn from practical application barriers and also be able to adapt the method based on interdisciplinary feedback.

In the next chapter, we will introduce the methodological approach and how it is applied in practice in more detail.

5 METHODOLOGICAL APPROACH AND OVERVIEW OF THE IMPACT AI METHOD MODULES

The Impact AI Method is a modular interdisciplinary mixed-method approach designed for external, ex-post impact assessments of running AI use cases, containing a technical survey and several empirical qualitative interviews, and their analysis and evaluation. While other impact assessments, for good reasons, argue that AI systems should be assessed before they are in use in order to foresee and prevent future damage (e.g. UNESCO's ethical impact assessment (2023)), our method is particularly designed to collect and reflect evidence of AI systems in use. With this approach, actual impacts in practice are assessed, while still allowing for corrections and improvements to the system, or, in a worst-case scenario, even the recommendation to overthink its implementation overall.

The current version of the Impact AI Method consists of 1) a preparation phase, 2) a technical survey based

on the AIMS templates covering the AI system, the model, and the dataset, 3) a semi-structured technical interview, followed by 4) three semi-structured "organizational" interviews, and additionally (tailored to each use case) 5) external impact interviews. After these steps in the data collection process, the final steps are 6) data analysis and 7) the assessment based on the 25 assessment criteria. An overview of these seven steps is given in the figure below:

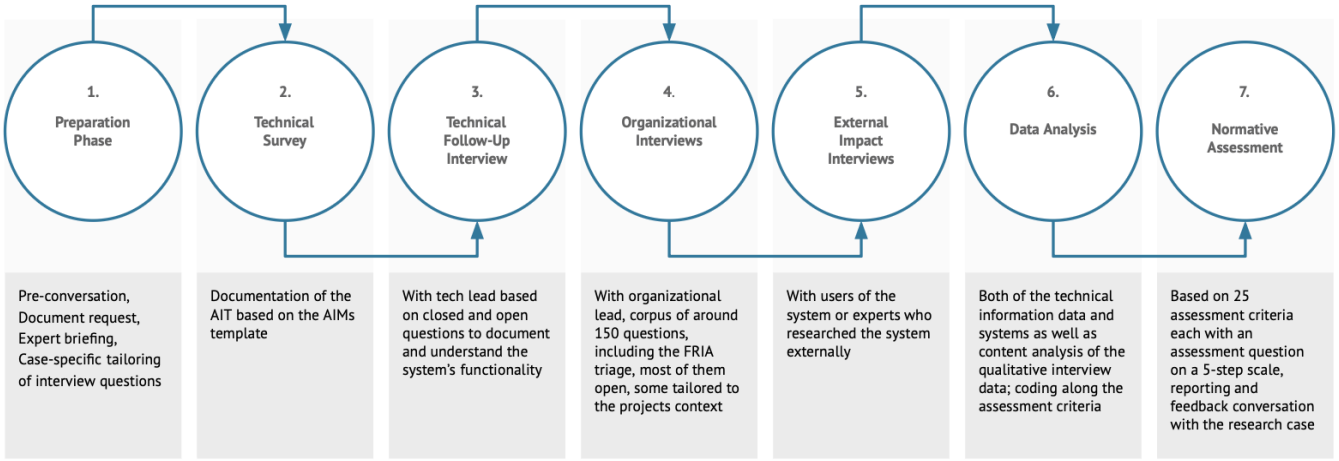


Fig. 1: The seven steps of the Impact AI Method.

5.1 The preparation phase

For the selection of AI use cases for our impact assessment, having a "pre-conversation" with the AI project owners proved very useful. This first exchange provides the opportunity to gather an initial understanding of the workings of the respective project teams and implementation of AI systems (while it is an open exchange, we have developed a "pre-questionnaire" which covers the most important topics to be clarified in this first conversation. The pre-questionnaire can be found here). It also enabled us to decide whether the project was suitable for the impact assessment. There were cases where AI projects proved unsuitable for the assessment as they were not yet in practice or it was clear that the project would end soon. In combination with this unrecorded first conversation, we conducted desk research on publicly available information on the AI use case. Common sources are the project website, project reports, research papers, news articles, and social media channels. These sources are included in our data analysis and help complement and validate the self-reported interview data from the projects. Since not all information is available through publicly accessible sources, we also directly request a list of documents from project owners. This includes, among others, detailed funding information, information on existing audits, risk or impact assessments (e.g. DPIA or FRIA), and project reports. We also request the technical documentation of the AI models and datasets. Additionally, prior to field research, we arrange briefing interviews with independent experts in the respective application domain, providing us with relevant contextual knowledge that helps us to better contextualize the specific AI use case. For instance, in a case researching an assistive AI tool for fact checkers, we consulted with an experienced researcher with domain knowledge on automation of fact-checking. Based on this information, as well as that provided by project owners and from our desk

research, we then tailor the interview guides for the impact assessment interviews to the specific AI project. Certain questions may be unsuitable for the case, while others may be particularly important or even require further exploration. Additionally, new project-specific questions may be added, e.g. on a specific aspect of the application domain, relevant to our assessment criteria.

5.2 The technical survey

The technical survey is based on the AIMS-templates, which contain three surveys, covering the AI system, the model, and the dataset individually (State et al., 2026). The technical survey is sent to the technical lead of the assessed project via email, to be filled in at any time within a certain agreed-upon period.
The modularization of the survey into three sub-surveys allows filling the survey also for complex AI systems (i.e. systems containing several models and datasets). Next to general information about the AI system, embedding, and hardware, it covers the following sustainability indicators: transparency, non-discrimination, privacy, and safety and security under social sustainability; complexity, energy use, and emissions, as well as resources under environmental sustainability; and model sharing, which includes model use and compatibility, under economic sustainability (an overview of all indicators within the template can be found in the appendix; the complete survey templates can be found here).

5.3 The technical follow-up interview

Based on the completed technical survey, we can prepare the follow-up technical interview with the technical lead of the AI use case. The goal of this interview is to ask clarifying questions regarding the information given in the survey or to address any questions that remained unanswered in the survey. Additionally, it contains some more in-depth questions in case the AI project uses an LLM or processes personal data. Overall, the follow-up technical interview is highly case-specific and closely tailored to the answers provided in the survey. The overall goal of the technical survey and this interview is to receive (at best) a full understanding of the AI system's architecture, used models and data, as well as hardware structure and the embedding of the AI system. A further goal is to be able to identify potentially critical aspects within this architecture and to understand potential risks of the use of the AI system (e.g. in regard to privacy or non-discrimination) and how they are technically mitigated (the interview guide for the technical follow-up interview can be found here).

5.4 The organizational interviews

We developed three questionnaires for interviews with the organizational lead of the AI project, overall consisting of around 95 to 128 questions (depending on how many follow-up questions are tailored to a use case): 1) the AI project impact and governance interview, 2) the AI project sustainability interview and 3) the AI project risk and mitigation interview. All interviews follow a semi-structured interview approach (Tisdell et al., 2025). Each interview guide is structured around a specific set of topics that summarize several indicators. Indicators themselves come with one or more interview questions, which we formulated to operationalize the indicator in question (the complete impact assessment methodology materials, including the interview guides, can be accessed here). At the same time, the semi-structured approach allows openness and flexibility to respond to the interviewees' answers and to address additional topics that emerge during the interview.

5.4.1 AI project impact and governance interview

The first interview focuses on the AI project's organization, impacts, and governance and includes reflections on the design choices of the AI system ("*What were relevant decisions about the architecture of your system (e.g., the languages in which the system works, the design of the interface?)*") and the decision processes behind these ("*Based on which arguments did you make these decisions and what was the decision-making process (e.g., stakeholder involvement)?*"). The interview also addresses the organizational goals and outcomes of the AI project, in particular the purpose of deploying the AI system. Also, questions regarding the participatory quality of the project and its orientation towards public interests are integrated. Further, we ask questions about impact achievements and outcomes for different impacted stakeholders, as well as respective evidence (e.g. "*What impacts (meaning outcomes and outputs) for each of these stakeholder groups do you see and which evidence can you provide for each?*") and potential harms. The interview also covers the necessity, proportionality, and alternatives of AI usage, and questions on the responsibility, accountability, and governance of the AI system are part of this interview. Table 1 displays the topics and indicators covered in this interview.

In our experience, both this interview and the AI project sustainability interview each require up to two hours. In some cases, a little more time was needed to clarify all relevant questions sufficiently.

Topic	Indicator
Organizational goals and outcomes of the project	Background AI system
	Decision process behind the AI system
	Goal and vision
	Connection to public interest
	Connection to sustainability
	Purpose of AI system
	Outcomes
	Outputs
Impact achievements and evidence	Impact achievements
	Stakeholder outcomes
	Strategy and obstacles
	Trade-offs in goals and impact
	Manifestations of good working conditions
	Relevance of impact achievements
Do no harm	Potential harms
	Harm management

	Purpose shift
	Safety and security
Necessity and proportionality	Proportionality of AI usage and role of AI
Responsibility, accountability and governance	Accountability
	Liability

Table 1: Overview of topics and indicators covered in the AI project impact and governance interview.

5.4.2 AI project sustainability interview

The second interview focuses on the three sustainability dimensions. When addressing social sustainability, we ask questions on human autonomy and oversight of the AI system, fairness and non-discrimination, transparency and explainability of the AI system, as well as questions on inclusion and participation of stakeholders (e.g. "*How much did stakeholder input influence project development and outcomes? Can you give an example of how your work was influenced by the perspective of affected stakeholders?*"). We cover environmental sustainability by asking questions on the environmental impact, in particular reflections on the impact in this regard (e.g. "*How do you see the environmental impact of AI systems in general and how do you reflect on your own system's impact in this regard?*" and "*Did your reflection lead to any type of action or change in your infrastructure decision?*"). Lastly, the part on economic sustainability includes questions on re-use and sharing of AI models and datasets, collaborations, and future perspectives of the AI project (e.g. "*What is your plan to keep the project available for the next 3-5 years? Please elaborate on concrete measures to maintain the project.*"). Table 2 displays the topics and indicators covered in this interview.

Topic	**Indicator**
Social sustainability	
Human autonomy and oversight	Human oversight
	Self-determined use
	AI literacy
Fairness and non-discrimination	Sovereignty and power relations
	Accessibility
Transparency and explainability	Transparency
Inclusion and participation	Intervention and participation
	Stakeholders and co-design
	Diversity
Environmental sustainability	

Environmental sustainability	Reflections on environmental impact
Economic sustainability	
Financial situation	Investment and maintenance costs for AI project
Openness and networks	Re-use and sharing
	Collaboration
Strategic outlook	Future perspective of AI project

Table 2: Overview of topics and indicators covered in the AI project sustainability interview.

5.4.3 AI project risk and mitigation interview

To assess potential negative impacts of AI systems, in particular impacts on fundamental rights, we have integrated a FRIA triage–a shorter version of a FRIA.

FRIA uses a scenario-based approach. Similar to this, we develop potential harm scenarios and outline potential fundamental rights risks that the AI project could cause, the severity and likelihood of the right violation, and possible prevention or mitigation measures. So far, we have done this in exchange with experts from the Center for Not-for-Profit Law.
The third interview covers the main part of the FRIA triage as the developed risk scenarios are addressed and discussed with the organizational lead of the AI project. We also inquire about the severity and likelihood of the rights violation and existing measures for prevention or mitigation (a FRIA triage template for scenario development can be found here). This, in turn, serves public interests by reflecting and understanding mitigation strategies for such harms appropriately. The exercise constitutes an important approach to handle potential harms of AI systems (Kieslich et al., 2026).

5.5 External impact interviews

To better understand intended as well as unintended effects on stakeholders and to determine whether these are positive or negative or if there are no effects at all, we seek to conduct interviews with either users of the relevant AI tool, other affected stakeholders, or experts who research user experience. These interviews are crucial, as they complement the findings from the interviews with the AI project and the self-reported data. The impact interviews are more openly designed and follow an individually prepared semi-structured interview guideline, tailored to the most relevant questions we would like to validate in regard to the impacts communicated to us by the core team of the running use case. Some examples of impact-oriented questions for this semi-structured interview are (interview guidelines for the external impact interview can be found here): On the integration and onboarding process of the AI system: *What did the integration process of the AI system look like, in particular, was there extensive training in regard to 1) potential risks and biases of the AI system or 2) the limitations of the AI system?* On the impacts, outcomes, and effects of the AI system: *How did the AI system affect you/ your working routines? How did you benefit from using the AI system?* On general feedback about the AI system: *Would you use the AI system again or recommend it to others?*

5.6 Data analysis

The collected data, including interview transcripts and documents, is analyzed using the software MaxQDA and following a qualitative content analysis. The assessment criteria (see Table 3) form the code framework, i.e. the categories used to label the interviews in order to systematically analyze the data.
Using an extractive and structured approach (Gläser & Laudel, 2010; Kuckartz, 2018), we can supplement these deductive codes by inductive codes emerging in the course of the analysis. The combination of such a systematic examination and openness to emerging codes ensures a comprehensive understanding of the impact of AI projects on sustainability and public interest at various levels.

5.7 Normative assessment

Our method also includes assessment criteria to further assess the results of the interviews and the survey in a next step. The results of these assessments are outlined and discussed in an impact assessment report, which includes a detailed account of our findings and recommendations for potential improvements for the AI project.

The objective of such an assessment along predefined criteria is twofold. First, it provides a learning experience for project owners, highlighting aspects within the organization and the AI system where improvements can be made or where performance is already strong. Second, the results of the assessment facilitate knowledge transfer to civil society and other sectors such as public administration, as they are presented in a clear and accessible way.

As already mentioned, the basis for the assessment criteria is "Principles for the Ethical Use of Artificial Intelligence in the United Nations System" (UN, 2022), which we extended with some criteria focusing on aspects of public interest and the three dimensions of sustainability. The criteria are aligned with five thematic clusters: impact, governance, social, environmental and economic sustainability (see Table 3, below).

We define the first criterion as a meta-level criterion, as its evaluation brings together several other criteria and gives answers to the research question: *Implications for people, society and the planet*. In addition to this meta-level criterion, three further criteria assess the **impact** of AI systems: the presence of clear *Objectives and their realisation* and whether they follow a coherent impact logic, evidence of *Impact achievements*, and the *Relevance of impact achievements* for the societal issue addressed. We further evaluate *Potential harms and mitigation,* and *Risk and harm management*. As aspects of public interest-oriented **governance** in the respective cluster, we assess *Necessity and proportionality* of AI usage, *Accountability*, *Decision processes and design*, and *Inclusion, participation, and co-design*. The **social sustainability** cluster includes *IT Security*, *Sovereignty and power relations*, *Accessibility*, and *Non-discrimination*, as well as *Data protection*, *Human oversight*, *Self-determined use*, and *AI literacy*. Regarding transparency, we assess the AI project's *Openness for validation and contestation*, *Technical explainability,* and *User transparency*. The **environmental sustainability** cluster covers the criteria *Complexity, energy and emissions,* and considerations about the use of *Natural resources*. Finally, we assess the existence of *Model and dataset sharing* and *Future perspective and collaboration* in the **economic sustainability** cluster.

For the assessment of the AI projects, for each of the 25 assessment criteria, we have developed an associated

assessment question. The data collection process we propose, including all its steps, including potential external impact interviews, should deliver the relevant information to answer the assessment questions. However, depending on the complexity of the use case, the openness of the respective interview partners, as well as the presence of additional evidence for this information, information can also be only partial, and assessment confidence can be low (which we then indicated). Each assessment question is also paired with one or more questions in the qualitative interview questionnaire or the technical survey, to make clear which statements the specific assessment question depends on (the complete catalogue of the assessment criteria and corresponding assessment questions can be found here). To facilitate answering the assessment questions and form a final judgment of each assessment criterion, we design and formulate additional helper questions. They support answering the assessment questions because they break down the assessment criteria into several aspects that are relevant for the final assessment. In this sense, they also help to maintain an overview of all relevant aspects for the specific assessment criteria and contribute to an objective and reproducible assessment.

The meta-level criterion is assessed qualitatively, and the assessment is written only. To answer the assessment question, the other 24 criteria are rated on a scale with the following five ratings: insufficient, weak, ambivalent, satisfactory, and exemplary.

Cluster	Assessment criteria	Corresponding assessment question
Meta-level	**Implications for people, society and the planet**	Which implications does the AI use case have for people, society and the planet? Does the overall impact of the AI project justify the claim to support public interest and/or sustainability?
Impact	**Objectives and Realisation**	Can goals, outcomes and outputs of the AI project be defined clearly and plausibly following a clear impact logic?
Impact	**Impact achievements**	Is there convincing evidence supporting the claim that the project has achieved envisioned outcomes?
Impact	**Relevance of impact achievements**	Are these outcomes relevant achievements in the given targeted problem area?
Impact	**Potential harm and mitigation**	Does the AI system in any way potentially cause or exacerbate harm, whether individual or collective, and including harm to social, cultural, economic, natural, and political environments and how is this harm prevented or mitigated?
Impact	**Risk and harm management**	Are there appropriate procedures to assess and monitor potential harms?
Governance	**Necessity and proportionality**	Is the use of an AI system, including the specific AI method(s) deployed, justified, necessary and proportionate in the given context and does it not exceed what is necessary and proportionate to achieve envisioned outcomes?
Governance	**Accountability**	Are there clear roles of responsibilities to ensure accountability for the AI system, its impacts and liabilities?

Governance	**Decision processes and design**	Is the design and decision-making process for the AI system reflecting principles of public interest and sustainability?
Governance	**Inclusion, participation and co-design**	When designing, deploying and using AI systems, does the project take an inclusive, interdisciplinary and participatory approach?
Social sustainability	**IT security**	Are there organizational responsibilities and measures to identify, address and mitigate IT security risks throughout the AI system lifecycle to prevent where possible, and/or limit any potential or actual harm?
Social sustainability	**Sovereignty and power relations**	Does the project contribute to shifting power in a democratic way (e.g. to civil society or marginalized groups) and support inclusive use and equal distribution of benefits and harms of the AI system?
Social sustainability	**Accessibility**	Does the project convincingly describe efforts or create measures to create accessibility?
Social sustainability	**Non-discrimination**	Does the project take appropriate measures to prevent bias, discrimination and stigmatization of any kind, in alignment with fundamental rights?
Social sustainability	**Data Protection**	Are the rights of data subjects respected, protected and promoted throughout the lifecycle of the AI system according to applicable laws?
Social sustainability	**Human oversight**	Does the design of the system ensure human oversight and human capability to oversee the overall activity of the AI system?
Social sustainability	**Self-determined use**	Do users make a self-determined choice when and how to use the system?
Social sustainability	**AI literacy**	Does the project describe efforts or concrete measures to improve AI literacy of people directly or indirectly impacted by the AI system?
Social sustainability	**Open for validation and contestation**	Does the project provide sufficient transparency and openness for evaluation?
Social sustainability	**Technical explainability**	Is technical explainability ensured, meaning that outputs by an AI system can be understood and traced by human beings?
Social sustainability	**User transparency**	Is there public information about the AI system that is understandable to the user of this system? Are individuals meaningfully informed about the AI system that they use, when a decision is informed by or made based on AI algorithms?
Environmental sustainability	**Complexity, energy and emissions**	Does the project sufficiently consider the complexity, energy and emissions of the AI system, including, but not limited to, measures to reduce these?

Environmental sustainability	**Hardware sustainability**	Does the project sufficiently consider the resource use and sustainability of their hardware?
Economic sustainability	**Model and dataset sharing**	Are the model and the dataset publicly shared, or are there good reasons not to share them, and did the project take measures to make it easy to use them?
Economic sustainability	**Future perspective and collaboration**	Are there efforts to establish collaboration and to sustain the project and the fulfilment of its objectives in the future?

Table 3: Overview of the 25 assessment criteria and their corresponding assessment questions, organised by cluster.

6 DISCUSSION

Applying our method to use cases in an early stage of its development was a helpful process to understand its strengths and weaknesses. We describe the most important learnings for improvements as well as open challenges for further development of the method.
First of all, our method produced valuable insights, which would not have been possible without the collaboration of the use cases and the data access we gained through documents, surveys, and the interviews. From our existing research, we derive several positive signs of the method's value.

We have derived **comprehensive assessment results for all criteria** based on our method. With the data gained from the interviews, assessment of all defined criteria was possible. Sometimes, however, this was possible with varying degrees of certainty, as some insights finally rely on self-descriptions of the project and cannot be conclusively validated (e.g. IT security processes or results from data protection risk assessments). The approach developed for data collection, analysis and assessment of the Impact AI Method enabled us to generate assessment results for the AI use case across various relevant topics. We regard the fact that we are able to derive meaningful assessments for each criterion–across a broad spectrum of topics–as proof that our method works in practice, while acknowledging that impact assessments, as always, capture two things: 1) the status quo of a system and do not allow findings over time, capturing the dynamic changes of a specific system. And 2) an interpretation of certain criteria (e.g. such a highly political concept as serving the public interest), which can to some extent be contested.
Further, the holistic assessment allowed us to provide **meaningful feedback to improve the assessed AI case and organization**. As we take a wide range of indicators into account–rather than focusing on a single thematic area–it enables us to assess the AI use case in many respects and identify areas for improvement with actionable recommendations for change.
The assessed **AI projects communicated that the feedback we delivered was helpful and instructive**. As one of our objectives is to create a learning experience for project owners, highlighting aspects within the organization and the AI system where improvements can be made or where performance is already strong, we communicate this feedback to the assessed AI use cases as part of the overall assessment process. We view the projects' positive response to the feedback as a sign of the value of our method.
In this regard, there have been requests for both the **future use of the impact assessment reports and the**

Impact AI Method itself for internal use by the assessed AI use cases. We see this interest in the method modules (e.g. the interview questions and assessment) as another sign of the method's value, not only in terms of its successful practical application but also in terms of the significance of the results to support internal reflection processes in a self-assessment.

As part of a **collaboration with an international organization** working in the field of sustainable development, we are testing the Impact AI Method for internal use to understand how it could support internal project development processes. This collaboration, which was proposed to us by this organisation itself, is another indicator of the practical usefulness and insightfulness of the results the method produces. We intend to develop more similar collaborations to apply the method in a programmatic manner and learn from these applications in organizational practices.

Finally, the results the method produces **contribute to the research** of so-called AI for good applications and the ongoing scholarly debate about societal impacts of AI.

Further, the in-depth case studies the method produces allow us to formulate several hypotheses for further research in this field: Firstly, we noticed from all case studies that user feedback is not always sufficiently integrated in the development flow of AI applications, even if specific societal impacts are the goal. We did, however, find that these external impact perspectives are meaningful for understanding the actual impacts and, in some cases, differ from the project owners' own impressions of their impact. Secondly, we can confirm from our case studies that a FRIA assessment is a very important cornerstone to reflect potential harms and improve mitigation strategies. Thirdly, all cases underlined the crucial importance of human roles (for oversight and accountability) which go far beyond any human-in-the-loop constellations. Instead, the practices embedding the AI system rather showed the incremental role of AI systems. While we could find how AI systems supported humans in specific tasks, human work, intentionality, and responsibility remained key to the actual impacts.

A productive tension stemming from the combination of public interest and sustainability, which enriched our findings, can be captured under the term 'global justice'. It addresses the issue of distribution of benefits from AI systems. In our view, this tension exemplifies a very relevant dilemma to the use of AI for good in general. Some of the assessed AI use cases relied to various degrees on LLMs from Big Tech (e.g. Google's Gemini or OpenAI's GPT models). While the production details of LLMs are nowhere near transparent, research and reporting have created awareness around the exploitative working conditions of data workers who were most likely essential to the production of both models (Miceli et al., 2024; Aapti Institute, 2026). The ambition to reflect the AI project's connection to public interest led us to the difficult moral question: Can an AI project be in the public interest and support sustainability if it is benefiting one public, while simultaneously hurting the rights of others (as data workers or affected communities from resource extraction in this scenario)? This question surfaces a recurring question around concepts of public interest: What is the scope of 'the public' concerned? Are we limiting our argument to the interest of a nation or global region, or do we need to consider the potentially harmful effects for other communities and regions as well? Sustainability, on the other hand, starts from a global perspective and thereby makes the evident argument that human rights and the rights of workers and the sustainable use of resources in global contexts are equally important in a consideration of positive or negative impacts. While this moral dilemma persists, the AI project's reflection of this issue was documented, and recommendations include the avoidance of such dilemmas and also the reliance on models of companies, which clearly prioritize profit maximization instead of public interests and sustainability. Despite the fact that they cannot change the

practices of big tech companies, we found these reflections relevant, as AI projects do, to some extent, have a moral obligation to reflect underlying conditions for their decision to use an AI system.

We also want to address the limitations when applying our method and explain how we handled them. For the selection of use cases, we are dealing with a strong selection bias, since projects need to agree to be assessed and be open to providing data for the documentation and interviews. This leads to the likely result of ethically or societally critical use cases not agreeing to be assessed, which is our supposition for some of the cases we reached out to but got a negative response for participation. Such projects might try to avoid public knowledge about details, problems, or even failures of their project, which would produce bad publicity and might affect future funding of their organization. Further, the cases we assessed do not include obvious and severe fundamental rights harms. This selection bias could lead to wrong assumptions about the field of AI for good projects, if it is assumed that we are assessing 'an average AI for good project'. However, we are well aware that the cases which do cooperate are probably closer to best practices than worst-case scenarios. Nevertheless, we did find issues that required deeper reflection and feedback, also in regard to potential fundamental rights impacts, even in these rather 'good' and low-risk cases, which illustrates the method's value. Finally, we are convinced that–even if our selection is limited–our results hold value for societal reflection and contribute evidence on the impacts of AI, as for now, the same level of depth in similar case studies is rare.
Another limitation results from self-reporting as a source for the assessment. While we put emphasis on including additional media reports, documentation, existing research, and interviews with external users or experts, many of our assessments still rely on self-descriptions of projects, which might be another source of bias. This limitation must be put into perspective, as without self-reported data, access to information and quality of an assessment would be even more limited. We make self-reporting transparent in our assessment and are cautious of its potential for distortion.

AI systems are never stand-alone products and only gain meaning within the socio-technical contexts in which they are embedded. The analysis and assessment of the impacts presented difficulty in clearly distinguishing between those impacts that stemmed from the AI system alone and those from the work of the project or even the organization overall. We also noticed that while our questions targeted the AI systems, answers often referred to overall experiences, efforts, or impacts of the project. On the one hand, we view this acknowledgment as a positive sign, as we believe that AI systems only function within their socio-technical environment, and that the long-term impact of AI—whether intended, unintended, positive or negative—only becomes apparent through interaction with the contexts they are embedded in. On the other hand, in our analysis we always tried to circle back and narrow our assessment to the direct effects of the AI system specifically. These outcomes can be smaller in scale, more immediate, and more incremental. While these outcomes are no less important than achieving long-term objectives, we found that they are less explicitly documented by the assessed use cases. To exemplify this: when assessing an assistive AI tool for fact checkers, we were interested in, on one hand, whether the project would achieve the envisioned outcome and impact (in this case, the aim was to contribute to better information and reduce misinformation or potentially harmful content). On the other hand, we were also interested in understanding the direct effect that using the AI tool would have on the user (in this case, the fact-checker). For example, we wanted to know how many working hours would be saved by using the

tool and how work processes would change. However, this was not assessed or reported specifically by the organisation.

With this reflection on the contributions and limitations of our methodological approach, we want to conclude and provide an outlook on how we plan to further develop and apply it.

7 CONCLUSION AND OUTLOOK

The core contribution of this paper is the detailed overview of the Impact AI Method, a structured and replicable approach for an impact assessment of individual running AI use cases aiming for a holistic assessment of implications for people, society and the planet.
While we have already iterated the methodology extensively, we foresee further iterations as a result of lessons in application. We hope to create societal impact by giving feedback to AI projects, providing evidence to the public, and bringing more nuanced findings on AI's impact for societal goals to academic and public debates. Overall, we hope to support a more informed, appropriate, and well-reflected embedding–or informed rejection–of the use of AI for societal goals and just sustainable development.
We see the assessment of impacts in regard to all dimensions of sustainability in depth as a contribution to not only wider transparency of the AI projects we research, but also as a good example for future legal or voluntary requirements, such as a report by the OECD on *Measuring the environmental impacts of artificial intelligence, compute and applications* proposed, where these impacts are framed as indirect impacts of AI (OECD, 2022). While we briefly mention the application of our approach, the analysis of our findings from the empirical assessments is beyond the scope of this paper and will be the focus of forthcoming publications. The method is first of all designed for researchers, but we aim to expand collaboration with other stakeholders such as NGOs and public bodies who develop and use AI, journalists, and also project owners of AI for good projects themselves. We work extensively to implement our method beyond the scope of this research project and make it available to stakeholders beyond academia. To this end, we actively share our method with other stakeholders in order to bring it into practice (the Impact AI Method is available here).

8 REFERENCES

Aapti Institute. (2026). *Engineered precarities: Algorithmic management in data work and content moderation*. Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ). https://aapti.in/blog/engineered-precarities-algorithmic-management-in-data-work-and-content-moderation/

AI + Planetary Justice Alliance. (2025). *AI Supply Chain Impact Framework. Evaluating AI's Environmental and Socio-Political Footprint.* https://aiplanetaryjustice.com/ai-supply-chain-impact-framework

Bozeman, B. (2007). *Public values and public interest: Counterbalancing economic individualism*. Georgetown University Press. http://www.jstor.org/stable/j.ctt2tt37c

Calvo, R. A., Peters, D., Cave, S. (2020). Advancing impact assessment for intelligent systems. *Nature Machine Intelligence, 2*, 89–91. https://doi.org/10.1038/s42256-020-0151-z

Chelimsky, E. (2014). Public-Interest Values and Program Sustainability: Some Implications for Evaluation Practice. *American Journal of Evaluation, 35*(4), 527–542. https://doi.org/10.1177/1098214014549068

Chui, M., Harryson, M., Manyika, J., Roberts, R., Chung, R., Nel, P., & van Heteren, A. (2018). *Notes from the AI frontier: Applying AI for social good*. McKinsey Global Institute. https://www.mckinsey.com/featured-insights/artificial-intelligence/applying-artificial-intelligence-for-social-good

Clarke, R. (2009). Privacy impact assessment: its origins and development. *Computer Law & Security Review, 25*, 123-135. https://doi.org/10.1016/j.clsr.2009.02.002

Cowls, J., Tsamados, A., Taddeo, M., & Floridi, L. (2021). A definition, benchmark and database of AI for social good initiatives. *Nature Machine Intelligence, 3*, 111-115. https://doi.org/10.1038/s42256-021-00296-0

Deutsches Institut für Normung. (2021). *DIN EN ISO 26000:2021-04 (D). Leitfaden zur gesellschaftlichen Verantwortung (ISO 26000:2010); Deutsche Fassung EN ISO 26000:2020.* DIN Media. https://dx.doi.org/10.31030/3210393

European Center for Not-for-Profit Law, & Danish Institute for Human Rights. (2025). *A Guide to Fundamental Rights Impact Assessments ("FRIA"). Under the EU Artificial Intelligence Act.* https://ecnl.org/publications/guide-fundamental-rights-impact-assessments-fria

European Commission: Directorate-General for Communications Networks, Content and Technology. (2020). *The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self assessment.* Publications Office. https://data.europa.eu/doi/10.2759/002360

European Union. (2025). *European approach to artificial intelligence*. European Commission. https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence

Fischer, R., Jakobs, M., Mücke, S., & Morik, K. (2022). A Unified Framework for Assessing Energy Efficiency of Machine Learning. In I. Koprinska, P. Mignone, R. Guidotti, S. Jaroszewicz, H. Fröning, F. Gullo, P. M. Ferreira, D. Roqueiro, G. Ceddia, S. Nowaczyk, J. Gama, R. Ribeiro, R. Gavaldà, E. Masciari, Z. Ras, E. Ritacco, F. Naretto, A. Theissler, P. Biecek, W. Verbeke, G. Schiele, F. Pernkopf, M. Blott, I. Bordino, I. Luciano Danesi, G. Ponti, L. Severini, A. Appice, G. Andresini, I. Medeiros, G. Graça, L. Cooper, N. Ghazaleh, J. Richiardi, D. Saldana, K. Sechidis, A. Canakoglu, S. Pido, P. Pinoli, A. Bifet, S. Pashami (Eds.), *Machine Learning and Principles and Practice of Knowledge Discovery in Databases. International Workshops of ECML PKDD 2022, Grenoble, France, September 19–23, 2022, Proceedings, Part I*. Springer, 39-54. https://doi.org/10.1007/978-3-031-23618-1_3

Flaksberg, D. (2024). *Designing with and Beyond Humans: Opportunities in Earth Centered Design Education.* 38th Hayes Advanced Research Forum, Ohio State University. https://kb.osu.edu/handle/1811/104250

Floridi, L., Cowls, J., King, T. C., & Taddeo, M. (2020). How to Design AI for Social Good: Seven Essential Factors. *Science and Engineering Ethics, 26*(3), 1771–1796. https://doi.org/10.1007/s11948-020-00213-5

Foth, M., Mann, M., Bedford, L., Fieuw, W., & Walters, R. (2021). Australia: A Capitalocentric Review of Technology for Sustainable Development: The Case for More-Than-Human Design. In A. Finlay (Ed.), *Global Information Society Watch 2020 Technology, the environment and a sustainable world: Responses from the global South*. Association for Progressive Communications (APC), 78-82. https://eprints.qut.edu.au/203186/

Gale, R. (2022). The war against sustainable development theory: Public interest as the ethical order for the 2030 Sustainable Development Goals. *Geographical Research, 60*(2), 347-357. https://doi.org/10.1111/1745-5871.12526

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., & Crawford, K. (2021). *Datasheets for Datasets*. arXiv. https://doi.org/10.48550/arXiv.1803.09010

Gläser, J., & Laudel, G. (2010). *Experteninterviews und qualitative Inhaltsanalyse als Instrumente rekonstruierender Untersuchungen. (4. ed.). Lehrbuch*. VS-Verlag, Wiesbaden.

Global Sustainability Standards Board. (2021). *GRI 1: Foundation 2021*. Global Reporting Initiative. https://www.globalreporting.org/how-to-use-the-gri-standards/gri-standards-english-language/?utm_source=google

Government Accountability Office. (2018). *Technology Assessment. Artificial Intelligence. Emerging Opportunities, Challenges, and Implications.* Government Accountability Office, Washington, D.C. GAO-18-142SP. https://www.gao.gov/assets/gao-18-142sp.pdf

Henderson, P., Hu, J., Romoff, J., Brunskill, E., Jurafsky, D., & Pineau, J. (2020). *Towards the Systematic Reporting of the Energy and Carbon Footprints of Machine Learning*. arXiv. http://arxiv.org/abs/2002.05651

High-Level Expert Group on Artificial Intelligence (2020). *The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self assessment*. European Commission. https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342

Hui, Z., & Karim, M. R. (2025). Pursuing climate justice through public interest litigation: Theories, practices, and prospects. *International Journal of Innovative Research and Scientific Studies, 8*(2), 4496–4510. https://doi.org/10.53894/ijirss.v8i2.6385

Hussein, M. A. & Sid Ahmed, H. M. (2025). The Role of Artificial Intelligence in Achieving the Sustainable Development Goals. *International Journal of Innovation and Knowledge Management in Middle East and North Africa, 10*, No. 1/2, 23-36. https://doi.org/10.47556/J.IJIKMMENA.10.1-2.2025.2

International Association for Impact Assessment. (n. d.). *Impact Assessment*. International Association for Impact Assessment (IAIA). https://iaia.org/impact-assessment/

IEEE. (2020). IEEE recommended practice for assessing the impact of autonomous and intelligent systems on human well-being. *IEEE Std 7010-2020*, 1-96. https://doi.org/10.1109/IEEESTD.2020.9084219

ISO/IEC 42005:2025 (2025): ISO/IEC 42005:2025: Information Technology — Artificial Intelligence (AI) — AI System Impact Assessment. https://www.iso.org/standard/42005

Jobin, A., Ienca, M., & Vayena, E. (2019). Artificial Intelligence: The Global Landscape of Ethics Guidelines. *Nature Machine Intelligence, 1*(2). https://doi.org/10.1038/s42256-019-0088-2

Kazim, E., & Koshiyama, A. (2020). Human Centric AI: A Comment on the IEEE's Ethically Aligned Design. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3575140

Kieslich, K., Helberger, N., & Diakopoulos, N. (2026) Scenario-based sociotechnical envisioning (SSE): An approach to enhance systemic risk assessments. *AI and Ethics, 6*(3), 255. https://doi.org/10.1007/s43681-026-01084-5

Kroll, C., Warchold, A., & Pradhan, P. (2019). Sustainable Development Goals (SDGs): Are we successful in turning trade-offs into synergies? *Palgrave Communications, 5*(1), 140. https://doi.org/10.1057/s41599-019-0335-5

Kuckartz, U. (2018). *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung. (4. Aufl.). Grundlagentexte Methoden*. Beltz Juventa, Weinheim, Basel.

Lewis, C. W. (2006). In pursuit of the public interest. *Publication Review, 66*(5), 694-701. https://www.jstor.org/stable/3843898

Lucaj, L., Loosley, A., Jonsson, H., & Gasser, U. (2025). *TechOps: Technical Documentation Templates for the AI Act*. arXiv. https://doi.org/10.48550/arXiv.2508.08804

Luccioni, A. S., & Hernandez-Garcia, A. (2023). *Counting Carbon: A Survey of Factors Influencing the Emissions of Machine Learning*. arXiv. https://doi.org/10.48550/arXiv.2302.08476

Luccioni, S., Jernite, Y., & Strubell, E. (2024). Power Hungry Processing: Watts Driving the Cost of AI Deployment? *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24)*, New York, Association for Computing Machinery. https://doi.org/10.1145/3630106.3658542

Martin, N., Friedewald, M., Schiering, I., Mester, B. A., Hallinan, D., & Jensen, M. (2020). *The Data Protection Impact Assessment according to Article 35 GDPR. A Practitioner's Manual.* Fraunhofer Verlag. https://publica-rest.fraunhofer.de/server/api/core/bitstreams/e6b91341-71f4-409b-8446-03432231a0d0/content

Miceli, M., Tubaro, P., Casilli, A. A., Bonniec, T. L., Wagner, C. S., & Sachenbacher, L. (2024). *Who Trains the Data for European Artificial Intelligence?* European Parliament, The Left, 1-40. https://hal.science/hal-04662589v1

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model Cards for Model Reporting. *FAT* '19: Proceedings of the Conference on Fairness, Accountability, and Transparency*, 220–229. https://doi.org/10.1145/3287560.3287596

Nasir, O., Javed, R. T., Gupta, S., Vinuesa, R., & Qadir, J. (2023). Artificial intelligence and sustainable development goals nexus via four vantage points. *Technology in Society, 72*, Article 102171. https://doi.org/10.1016/j.techsoc.2022.102171

Munn, L. (2024). The five tests: Designing and evaluating AI according to indigenous Māori principles. *AI & SOCIETY, 39*(4), 1673–1681. https://doi.org/10.1007/s00146-023-01636-x

Organisation for Economic Co-operation and Development. (2022). *Measuring the environmental impacts of artificial intelligence compute and applications: The AI footprint.* OECD Digital Economy Papers Nr. 341. https://doi.org/10.1787/7babf571-en

Organisation for Economic Co-operation and Development. (2024). *Recommendation of the Council on Artificial Intelligence.* OECD/LEGAL/0449. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

Palomares, I., Martínez-Cámara, E., Montes, R., García-Moral, P., Chiachio, M., Chiachio, J., Alonso, S., Melero, F. J., Molina, D., Fernández, B., Moral, C., Marchena, R., Pérez de Vargas, J., & Herrera, F. (2021). A panoramic view and swot analysis of artificial intelligence for achieving the sustainable development goals by 2030: progress and prospects. *Applied Intelligence, 51*, 6497–6527. https://doi.org/10.1007/s10489-021-02264-y

Raji, I. D., Smart, A., White, R.N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. *Proceedings of the 2020 conference on fairness, accountability, and transparency. ACM*, 33-44. https://doi.org/10.1145/3351095.3372873

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Official Journal of the European Union*, L 2024/1689. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

Reisman, D., Schultz, J., Crawford, K., & Whittaker, M. (2018). *Algorithmic impact assessments: a practical framework for public agency accountability.* AI Now Institute. https://ainowinstitute.org/publications/algorithmic-impact-assessments-report-2

Rohde, F., Wagner, J., Meyer, A., Reinhard, P., Voss, M., Petschow, U., & Mollen, A. (2024). Broadening the Perspective for Sustainable Artificial Intelligence: Sustainability Criteria and Indicators for Artificial Intelligence Systems. *Current Opinion in Environmental Sustainability, 66*, 101411. https://doi.org/10.1016/j.cosust.2023.101411

Ryan, M. (2025). We're only human after all: A critique of human-centred AI. *AI & SOCIETY, 40*(3), 1303–1319. https://doi.org/10.1007/s00146-024-01976-2

Sætra, H. S. (2021). AI in context and the sustainable development goals: Factoring in the unsustainability of the sociotechnical system. *Sustainability, 13*(4), 1738. https://doi.org/10.3390/su13041738

Selgas-Cors, M. (2025). Sociotechnical Transformation: A Systematic Review on the Impact of Artificial Intelligence on Society and Organizations. *FinTech and Sustainable Innovation*, 1-16. https://doi.org/10.47852/bonviewFSI52026076

Stahl, B. C., Antoniou, J., Bhalla, N., Brooks, L., Jansen, P., Lindqvist, B., Kirichenko, A., Marchal, S., Rodrigues, R., Santiago, N., Warso, Z., & Wright, D. (2023). A systematic review of artificial intelligence impact assessments. *Artificial Intelligence Review, 56*(11), 12799–12831. https://doi.org/10.1007/s10462-023-10420-8

State, L., Züger, T., & Winter, L. (2026): The AIMS-Templates: Documenting the multidimensional sustainability of AI systems. *Proceedings of the 2026 International Conference on Information Technology for Social Good (GoodIT '26)*. Association for Computing Machinery. https://doi.org/10.1145/3794786.3830766

Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and Policy Considerations for Deep Learning in NLP. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 3645–3650. https://doi.org/10.18653/v1/P19-1355

Tisdell, E. J., Merriam, S. B., & Stuckey, H. (2025). *Qualitative Research: A Guide to Design and Implementation. (5. ed.)*. Jossey-Bass.

Tomašev, N., Cornebise, J., Hutter, F., Mohamed, S., Picciariello, A., Connelly, B., Belgrave, D., Ezer, D., Haert, F., Mugisha, F., Abila, G., Arai, H., Almiraat, H., Proskurnia, J., Snyder, K., Otake, M., Othman, M., Glasmachers, T., Wever, W., & Clopath, C. (2020). AI for social good: unlocking the opportunity for positive impact. *Nature Communications, 11*, Article 2468. https://doi.org/10.1038/s41467-020-15871-z

United Nations Educational, Scientific and Cultural Organization. (2023). *Ethical Impact Assessment: A Tool of the Recommendation on the Ethics of Artificial Intelligence.* https://doi.org/10.54678/YTSA7796

United Nations. (1948). *Universal Declaration of Human Rights*. https://www.un.org/en/about-us/universal-declaration-of-human-rights

United Nations. (2009). *Guidance Note of the Secretary-General on Democracy.* https://www.un.org/democracyfund/sites/default/files/un_sg_guidance_note_on_democracy.pdf

United Nations, General Assembly. (2015). *Transforming our world: the 2030 Agenda for Sustainable Development.* A/RES/70/1. https://digitallibrary.un.org/record/3923923?v=pdf#files

United Nations. (2022). *Principles for the Ethical Use of Artificial Intelligence in the United Nations System*. United Nations System, Chief of Executives Board for Coordination (CEB), High-level Committee on Programmes (HLCP). https://unsceb.org/principles-ethical-use-artificial-intelligence-united-nations-system

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S. D., Tegmark, M., & Nerini, F. F. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. *Nature Communications, 11*, 1–10. https://doi.org/10.1038/s41467-019-14108-y

Weidner, H. (2002). *Gemeinwohl und Nachhaltigkeit—Ein prekäres Verhältnis*. Wissenschaftszentrum Berlin für Sozialforschung. https://bibliothek.wzb.eu/pdf/2002/ii02-303.pdf

World Commission on Environment and Development (1987). *Our Common Future: Report of the World Commission on Environment and Development.* https://sustainabledevelopment.un.org/content/documents/5987our-common-future.pdf

Züger, T., Fassbender, J., Kuper, F., Nenno, S., Katzy-Reinshagen, A., & Kühnlein, I. (2022). *Civic Coding. Grundlagen und empirische Einblicke zur Unterstützung gemeinwohlorientierter KI.* Civic Coding Initiative. https://www.conpolicy.de/aktuell/civic-coding-grundlagen-und-empirische-einblicke-zur-unterstuetzung-gemeinwohlorientierter-ki/

Züger, T., & Asghari, H. (2023). AI for the public. How public interest theory shifts the discourse on AI. *AI & SOCIETY,38* (2), 815–828. https://doi.org/10.1007/s00146-022-01480-5

9 APPENDIX

Overview of the AI system template	
Topic	**Indicator**
Description and development	Description
	Main development steps
Embedding	Description
	Technical aspects
	Additional data collection
Hardware	Description
	Updates
Social sustainability	
Transparency	Documentation
	Audit trails and other logs
Safety and security	Evaluation
	Mitigation measures
Environmental sustainability	
Complexity, energy use and emissions	Usage information
Natural resources	Decommissioning plans
	Certificates

Overview of the AI model template	
Topic	**Indicator**
Date and version	Date and version
Description and details	Type of model
	Architecture
	Task of model
	Data
	Languages and libraries
Training	Details model training
	Retraining

Performance	Metrics and benchmarks
	Factors
	Uncertainty and variability
Social sustainability	
Transparency	Interpretability
	Explainability methods
Non-discrimination	Evaluation (fairness metrics)
	Methods for model debiasing
Privacy	Evaluation
	Measures to improve privacy
Environmental sustainability	
Complexity, energy use and emissions	Complexity measures
	Tracking or estimation of emissions
	Measures to reduce complexity
Economic sustainability	
Sharing	Public availability
	Citation and license
	Contact person and maintaining
Usage	Use, incl. out-of-scope use
Compatibility	Compatibility
Additional harm	
Harms	Additional harms and mitigation

Overview of the dataset template

Topic	**Indicator**
Collection	General aspects and sampling
	Labeling
	Timeframe
	Purpose
Description	Data modality

	Description features
	Meaning of instances
	Label / target
	Size
	Relations and external sources
	Missing information, errors, confidentiality or offensiveness
Processing	Preprocessing
	Splits
Maintenance	Updates
	Storage
Social sustainability	
Non-discrimination	Representativeness
	Methods for rebalancing
	Sensitive features or proxies
	Measures
Privacy	Evaluation
	Measures
Environmental sustainability	
Complexity, energy use and emissions	Complexity measures
	Reduction of complexity
Economic sustainability	
Sharing	Public availability
	Citation and license
	Contact person and maintaining
Usage	Use, incl. Unwanted use
Compatibility	Compatibility
Additional harm	
Harms	Additional harms and mitigation

Table 4: Overview of topics and indicators covered in three templates (on the AI system, the AI model and dataset) constituting the technical survey.