# Sustainability for Artificial Intelligence Products and Services – Initial How-to for IT Practitioners

Dominic Lammert
Furtwangen University and LUT University
dominic.lammert@hs-furtwangen.de

Larry Abdullai
LUT University
larry.abdullai@lut.fi

Stefanie Betz
Furtwangen University and LUT University
stefanie.betz@hs-furtwangen.de

Jari Porras
LUT University
jari.porras@lut.fi

## Abstract

*Year after year, software engineers celebrate new achievements in the field of AI. At the same time, the question about the impacts of AI on society remains insufficiently answered in terms of a comprehensive technology assessment. This article aims to provide software practitioners with a theoretically grounded and practically tested approach that enables an initial understanding of the potential multidimensional impacts. Subsequently, the results form the basis for discussions on AI software requirements. The approach is based on the Sustainability Awareness Framework (SusAF) and Participatory Design. We conducted three workshops on different AI topics: 1. Autonomous Driving, 2. Music Composition, and 3. Memory Avatars. Based on the results of the workshops we conclude that a two-level approach should be adopted: First, a broad one that includes a diverse selection of stakeholders and overall impact analysis. Then, in a second step, specific approaches narrowing down the stakeholders and focusing on one or few impact areas.*

**Keywords:** Sustainable Artificial Intelligence, Software Sustainability, Requirements Engineering, Software Engineering, Software Development

## 1. Introduction

AI is expected to have an immense impact on our lives through possibilities such as autonomous driving, better healthcare services, big data analytics, and even employment opportunities. On the other hand, it could also, become the worst event in the history of humanity (Vöneky, 2020). In any case, it is hard to deny that AI is changing and will continue to change our lives in intended and unintended ways (Rahwan et al., 2019). Therefore, the positive changes brought by AI should be contrasted with the adverse effects of this technology.

Van Wynsberghe proposes the following definition of sustainable AI: ”Sustainable AI is a movement to promote change throughout the lifecycle of AI products (i.e. idea generation, training, tuning, implementation, governance) towards greater environmental integrity and social justice. Sustainable AI thus focuses not only on AI applications but on the entire AI socio-technical system.” Furthermore, the researcher suggests that in order to be sustainable, AI places sustainable development at the core of its development ”with the three associated tensions between AI innovation and equitable resource distribution, inter-and intra-generational equity, and between the environment, society and the economy.” (Wynsberghe, 2021). The development of AI software is thus a dual-task. It has to be approached from both technical and social justice points of view. Technical considerations refer to the performance of the AI system. They can be described as functional properties that can be examined using metrics from the field of machine learning, such as accuracy or precision. On the contrary, the social justice considerations are undoubtedly no less complex. These include ensuring principles such as transparency, interpretability, and fairness. We are confronted with non-functional properties that prove to be much more complex because we cannot rely on standardized metrics and procedures from the field of machine learning.

Therefore, inter- and transdisciplinary research is needed to develop and implement suitable testing strategies. The diversity of possible use cases for AI is too great for a single universal solution to suffice. In addition, complex systems are usually not static but are subject to constant change. Hence, it must be continuously iterative, improved and optimized.

URI: https://hdl.handle.net/10125/103428
978-0-9981331-6-4

Numerous studies and frameworks focus on a single selected impact within an impact dimension of AI software. For example, one of the best known and among the most studied challenges in the social dimension is fairness (Agarwal et al., 2022; Angell et al., 2018; Brun & Meliou, 2018; Chouldechova & Roth, 2020; Sharma et al., 2019). Scientists repeatedly point out in their conclusions that further research on these areas are relevant. There is yet or will never be a one-size-fits-all solution. In this article, we would like to take a bird's eye view of the impacts of AI software, not focusing our attention on a single selected impact, but looking at the "big picture". Vöneky explains that AI is a complex subject to understand completely (Vöneky, 2020). Similar argument applies to the multidimensional and multilayered impacts of AI.

Nonetheless, despite having a plethora of academic discourse and guidelines concerning, for instance, AI ethics (Berendt, 2019), yet, research conducted by the Pew Research Center revealed that experts doubt that ethical AI or for that matter sustainability issues will be at the center of AI design in the next decade (Rainie et al., 2021). In the report, the researchers posit that developers and designers of AI are primarily focused on profit and social control at the expense of the possible consequences. In other words, for many AI designers and developers, there is an attitude of being the first to innovate and rectify the damages late. The problem with such an attitude is that some of the consequences may not be reparable after the damage has been done. According to Berendt, there are four characteristics of AI practices that may account for reasons why the notion of "common good" or reflecting on the potential adverse effects of AI are not considered during the design face: the problem-solving and approach of the AI engineer or developer, the inclusion of diverse stakeholders, the role of knowledge, and the awareness of side effects and dynamics (Berendt, 2019).

Software practitioners, in particular, lack the knowledge and methods to consider AI software impacts on software requirements (Galaz et al., 2021; Khakurel et al., 2018; Wynsberghe, 2021). Perhaps, having a simple but encompassing framework which engages both AI engineers, developers, researchers, policymakers, users and other stakeholders on the sustainability issues that could potentially be affected as a result of a technology, product or service could be the starting point for AI engineers to consider sustainability as a requirement during AI design face. Thus, we pose the following research question: **What does software practitioners need to pay attention to in terms of sustainability impacts when developing AI software?**

To answer our research question, we conducted three workshops using the *Sustainability Awareness Framework (SusAF)* according to the principle of *Participatory Design*. The SusAF is a tool developed by an international group of researchers that aims to raise awareness of the relationship between software and social, individual, environmental, economic, and technical sustainability, as well as their potential immediate, enabling, and structural impacts.We brought together different stakeholders to discuss AI software, identified potential impacts, and thus provide guidance on how such systems should and should not be designed. We conducted the workshops on three completely different topics: Autonomous Driving, Music Composition and Memory Avatars.

In all three workshops, several implications emerged that had not been considered beforehand. As a result, it is essential to take a two-level approach when designing AI. At first, a broad one that includes a diverse selection of stakeholders and a multidimensional impact analysis at first. Then, in a second step, specific approaches should be used, narrowing down the stakeholders and focusing on one or a few selected impacts.

## 2. Background

Khakurel et al. recognize that AI companies are showing an increased interest in joining the AI trend. However, it is unclear what social, environmental, and economic impacts this will have (Khakurel et al., 2018). Meantime, we are confronted with the fact that trust in AI is rather low among the general population. The study "Trust in Artificial Intelligence - A five country study" (Gillespie et al., 2021), based on surveys in the USA, Canada, Germany, Great Britain, and Australia indicates that "most citizens being unwilling or ambivalent about trusting AI in healthcare (63 percent) and HR (77 percent)." Therefore, we must enhance trust in AI products and services. Since the foundation of a system are its requirements, we suggest starting with the assessment of the possible impacts during the requirements engineering phase.

### 2.1. Requirements Engineering for AI

What makes AI Products and Services so unique in their effects that they must be examined separately from other technologies? The answer to this can be found in its socio-technical impacts.

*Technical side:* In an article on the interplay of requirements, technology, and AI, Kostova et al. conclude that their analysis raises more questions than answers (Kostova et al., 2020). Ahmad et al. answer this question in their SLR study, "What is up with Requirements Engineering for Artificial Intelligence

Systems" by stating that the development process of AI systems differ from traditional approaches (Ahmad et al., 2021; Sculley et al., 2015). The authors of the SLR study recommends that requirements engineers bridge with data scientists and machine learning specialists. They refer to Amershi et al. (Amershi et al., 2019), who again point out that "both data scientists and software engineers should improve their knowledge and understanding of the issues that arise from incorporating AI into most software projects and learn to work together."

*Social justice side:* Software engineers focus too one-sidedly on technology - "artificial systems with clear boundaries and identifiable parts and connections, modules and dependencies (Becker et al., 2016)" - while other systems, such as social, environmental, and economic consequences are not sufficiently considered. At this point, a second SLR study by Ahmad should be mentioned, which allows for a "human-centric" approach due to the considerable new challenges in Requirements Engineering (Ahmad et al., 2021). Ahmad explains: "Human-centric approaches involve providing systems that are interpretable, explainable, transparent, secure and fair". Here the author refers to Fagbola and Thakur, who want the aspect of multidimensional impacts to be understood as an open problem (Fagbola & Thakur, 2019). In Ahmad's opinion, there is little research on Requirements Engineering techniques for building AI systems.

When requirements engineers plan the use of AI software, they must not neglect the unintended and unforeseen impacts of AI systems. Add to this the fact that these impacts may not be foreseeable. The system will probably need to be iteratively readjusted even after market entry. At this point, reference should be made to the report "The Ethical Skills We Are Not Teaching" (Suárez & Varona, 2021) of Suárez and Varona. The authors conducted a textual analysis of 503 courses on non-functional issues of AI at 66 universities in 16 states and conclude that instructors are not training their students in ethical skills. Bogina et al. come to a similar conclusion and recommend that the need for such education must be met to meet the challenges of AI impacts (Bogina et al., 2021).

### 2.2. Sustainability Frameworks in AI

Establishing social justice development of software has become increasingly important in recent years. One of the organizations that comprehensively addresses the software practitioner's endeavor to integrate social, environmental, and economic issues in terms of ethics, morality, and sustainability is AlgorithmWatch: "AlgorithmWatch is a non-profit research and advocacy organization that is committed to watching, unpacking and analyzing automated decision-making (ADM) systems and their impact on society."[1] Today, the organization lists over 160 tools, which we looked at in preparation for this article. 66 tools are directly or indirectly focused on our endeavour, and we subjected them to closer analysis.

The frameworks can be roughly categorized: First, is generally focused on multiple and diverse areas. They take a bird's eye view of the possible impacts of the software systems and are thus suitable for identifying and discussing them. The *(SustAIn)* and the *(SusAF)* should be mentioned here. In addition, it is noticeable that these frameworks are aimed at software practitioners. Second, it addresses specific, selected issues, for example, bias *(The Imperial Machines Project)*, fairness *(Fairness Aware Ranking)* and privacy *(VBRE)*. In most cases, they are aimed at software developers. The third is the development of technical tools for checking algorithms. Most of these tools, including *AI-Fairness360* and *Fairlearn*, for example, focus on analyzing biases in data sets. In most cases, they are aimed at data scientists.

There are a variety of arguments that led us to apply the SusAF in our workshops. It helps software practitioners engage in conversation with different stakeholders. Thus it is a participatory approach. Participants look together for interactions between software and five dimensions: social, individual, environmental, economic, and technical. Additionally, SuSAF enables participants to identify impacts over time, different dimensions and time layers can be analyzed. In addition, the SusAF has a straightforward procedure designed as a workshop; thus, it is easy to apply.

## 3. Empirical study

The study utilized data from three different workshops to gain insight into the possible sustainability impacts of AI-based software. To achieve this, we used Participatory Design and the SusAF to identify the effects of AI. The subsequent section discusses the processes and methods adopted to carry out our explorative mixed-method study.

### 3.1. Methodological background

**3.1.1. Participatory Design (PD)** The overarching aim of PD is to enable diverse stakeholders to interact

[1] https://inventory.algorithmwatch.org/

and actively participate in contributing different layers of tacit knowledge and expertise for value co-creation and co-design. In the context of our study, this means the engagement of project managers, developers, potential customers, researchers, and other stakeholders to examine various aspects of sustainability issues in the design and development of AI. This methodology enabled us to make recommendations for policies, actions, industry, and society (Grunwald, 2020). According to Simonsen and Robertsen (Simonsen & Robertson, 2012), there are four stages in PD, namely Requirement Analysis, Analysis and Design, Implementation, and Test. Nonetheless, this paper only focuses on the Requirement Analysis as the first step into this new research phenomenon. Furthermore, the focus of PD lies on three types of sources (i.e., stakeholders, documents, and systems) to ensure user-centered design.

**3.1.2. Sustainability Awareness Framework (SusAF)** The SusAF is a sustainability impact awareness tool that provides a set of questions, a visualization tool, guidelines, and templates that help software practitioners to identify and discuss potential sustainability impacts of their AI on people, society, and IT systems. Becker et al. have shown that SusAF helps to identify potential effects and chains of effects of a sociotechnical system and start a conversation about its impacts (Becker et al., 2016). Companies often are aware of the direct impact of their IT products and services but SusAF challenges designers and businesses to reflect beyond and be aware of the systemic chain of effects of their IT systems. Hence, participants are supported by scenarios to consider not only the immediate characteristics and impacts of their product or service but also their medium-to-long-term interconnected chain of effects (see tables 1 and 2). Using the Sustainability Awareness Diagram (SusAD), a radar chart, we can to map out the positive and negative chains of effects (Duboc et al., 2019), that AI software could potentially have based on the five sustainability dimensions (see figures 1, 2 and 3). Finally, as mentioned earlier, the SusAF process is straightforward and already designed as a workshop.

## 3.2. Workshop design

**3.2.1. Planning phase** The study is empirically supported by three workshops, with each addressing different topics: AI for Autonomous Driving, AI for Music Composition, and AI for Memory Avatars. Following (Simonsen & Robertson, 2012), we gathered three types of sources: Stakeholders, Documents, and Systems in operation for our PD. *Stakeholders* as a source means the recognition of the process requires a heterogeneous composition of participants. The participants must feel represented in a balanced way. The appointment of mediators could promote a balanced agreement and fair compliance with the rules of procedure. We included the various stakeholders in the participant composition. *Documents* contain information from which requirements can be derived. These can be experience reports, legal texts, standards, ethical value discussions, error reports on suitable alternative systems, etc. *System in operation* by testing and analyzing predecessor and/or competitor systems may result in new or modified requirements. Below, we describe how each of the three sources was used in our study. The first step was a schematic stakeholder identification. In particular, we discussed these questions: Who is considered to be affected by the project, and what exact processes are affected by the proposed project? Initially, we came up with a list of stakeholders, which we had to evaluate, prioritize and map (see section 4. Results).

This study is supported empirically by both primary (workshop) and secondary data (literature). We first reviewed and analyzed scientific and grey documents to understand the state-of-the-art of the phenomenon, its challenges, and future promise. Armed with this background, we were able to explore and select areas to carry out our sustainability awareness workshops and identified some possible sustainability impacts of AI systems.

The schedule of workshops was as follows:

1. **Warm-up:** One of the authors (i.e., facilitator) presents the SusAF and the purpose of the workshop. Another short presentation followed this by the project manager about their AI product or service and their initial perception about the sustainability impact of their AI product or service which participants were allowed to contribute.

2. **SusAF:** Next was to use PD to engage, challenge, and provoke participants' thoughts about the sustainability impacts of the topic under discussion with the sets of questions from SusAF as a guide.

3. **Discussion:** Finally, participants together with one of the authors discussed the answers and summarized the elements of the SusAF that were most important to them and clarified any ambiguities. Together, the three parts of each workshop lasted about three hours.

Table 1. The three types of effects based on the SusAF (Penzenstadler et al., 2020)

| Effect | Description |
|---|---|
| Immediate | ”are direct effects of the production, operation, use and disposal of socio-technical systems.” |
| Enabling | ”of operation and use of a system include any change enabled or induced by the system.” |
| Structural | ”represent structural changes caused by the ongoing operation and use of the socio-technical system.” |

Table 2. The five dimensions of sustainability based on the SusAF (Penzenstadler et al., 2020)

| Dimension | Description |
|---|---|
| Social | ”covers the relationships between individuals and groups.” |
| Individual | ”covers the ability of individuals to flourish, exercise their rights, and develop freely.” |
| Environmental | ”covers the use and stewardship of natural resources.” |
| Economic | ”covers the financial aspects and business value.” |
| Technical | ”cover the technical system’s ability to accommodate changes.” |

4. **Survey:** At the end, each participant received a survey questionnaire regarding the content and the structure of the workshop.

**3.2.2. Data collection** The participatory workshops were divided into three main sessions. In the first session, we provided an introduction to the SusAF workbook and explained the online whiteboard workstation to the participants. The participants were sent links to the whiteboard a day before the workshop to familiarise themselves with the tools for a smooth workshop. All technical issues regarding the use of the whiteboard were cleared in this session. The second session was the engagement part where participants were actively engaged in the brainstorming of the perceived sustainability impact of the topic under discussion. The participants wrote their responses on the Miro board from which the SusAD was developed. The final session was the feedback part, where participants gave verbal comments about the workshop and completed a survey. The survey included personal questions about the characteristics of the participants, such as gender, industry and position, age, and years of work experience, to a broader perspective about sustainability awareness, the importance of the workshop, and the engagement method (see section 4.4).

**3.2.3. Data analysis** After the workshop, the authors individually collected and summarized the data from the Miro board into SusAD. We then compared our findings and where there were discrepancies, we revisited the original data on the Miro board to reconcile any differences. Finally, after resolving all ambiguities, we had three different SusADs showing the potential sustainability impacts and chain-of-effects for each of the topics discussed in the workshop.

# 4. Results

We made sure to involve suitable stakeholders in each workshop:

- **Autonomous Driving:** *Age:* between 22-41 (Ø 29); *Gender:* 2 female and 7 male; *professions:* mechanical engineer, urban planner, UX Designer, 6 IT students; *Working experience:* 2 to 14 years (Ø 5 years)
- **Music Composition:** *Age:* 25-38 (Ø 34); *Gender:* female: 1, male: 4; *professions:* 4 Researchers (Music Science and Software Engineering) and IT student; *Working experience:* 1 to 2 years (Ø 1.4 years)
- **Memory Avatars:** *Age:* 18-69 (Ø 30); *Gender:* female: 20, male: 10; *professions:* University: 10 Social Sciences, 8 Computer Science, 3 Theology, 2 Educational Science, Environmental Science, Media Science, Industry and church: Account Manager, Grief counselor and content manager, Licensing manager, Pastor, Theologian (also a supervisor and personal coach); *Working experience:* 0 to 38 years (Ø 5.4 years)

Each participant was offered a one-on-one or group session in advance that included an introduction to the technical basics of AI. This allowed us to ensure that even those participants who were not previously familiar with AI had a basic understanding of the subject matter and were familiar with the common technical vocabulary.

At the outset, we must point out that the findings in Sections 4.1 through 4.3 are primarily intended to understand the potential contribution of this research. Nevertheless, some arguments need to be verified and supported by studies. Thus, the three diagrams are by no means intended to represent the state of the science.

### 4.1. Autonomous Driving

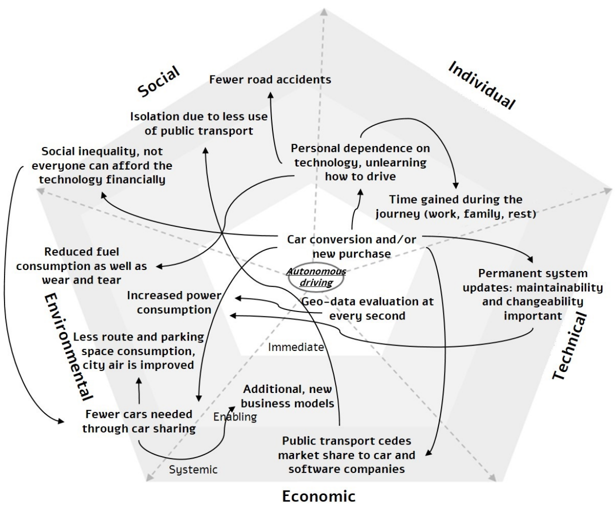


Figure 1. SusAD: Autonomous Driving

At the beginning of the first analysis, it should be noted that in comparison to the other two analyses, fundamental contradictions arose between the stakeholders in the case of Autonomous Driving. One of the first impacts that arose was the fact that drivers would have to retrofit their car or buy a new one (immediate effect in the individual dimension). This impact is linked to a positive and a negative impact chain. On the one hand, the stakeholders expect savings because fewer cars are needed through car-sharing (systemic effect in the environmental dimension). A circumstance that should lead to further business models (enabling effect in the economic dimension). Another positive impact that appears in this chain of effects is the reduced consumption of space on the streets and car parks, which in turn benefits the city air (systemic effect in the environmental dimension). However, the retrofitting and purchase of new cars would also result in public transport losing market share to car and software companies (systemic effect in the economic dimension). This in turn would lead to social isolation within society, which would be the result of less use of public transport (systemic effect in the social dimension).

Another positive impact is expected on the individual dimension. Although the enabling effects show that the ability to drive a car is lost over time, in return one gains time that can be used more sensibly for family, work and rest.

In the environmental dimension, there is disagreement about whether CO2 consumption will rise or fall as a result of Autonomous Driving. Permanent system updates and geo-data evaluation (technical dimension) lead to an increase in energy consumption. On the other hand, the previously mentioned decrease in the number of cars on the road and lower fuel consumption. In addition, the wear and tear on the car should also decrease.

### 4.2. Music Composition

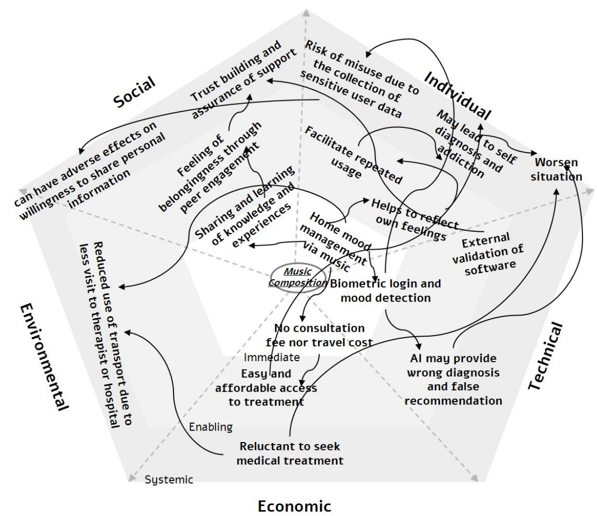


Figure 2. SusAD: Music Composition

The chain of effects (as shown in figure 2) makes AI-based music selection interesting. While the service results in many direct impacts, it also has enabling and unintended systemic consequences. For instance, participants mentioned that the primary stakeholder, thus the individual user can directly and positively manage his/her mood at home without visiting the hospital nor the psychologist. However, this positive impact creates self-awareness of an individual's immediate environment and therefore enables repeated listening to the recommended music. Overtime, this may negatively lead to self-diagnosis and addiction and also, worsen the mental health situation of the user.

From the economic perspective, mood management in the comfort of an individual's home through music recommended by AI positively eliminates medical consultation fees and travel costs as a direct impact. As a result, it means easy and affordable access to treatment. However, this may cause users to feel reluctant to seek medical treatment in the long-run. Although this has enabling effect on the environment by less usage of transport, the systemic effect is that it gradually deteriorates the health condition of the individual user. Furthermore, to access the app, one of the technical requirements is personal login and mood recognition.

Although, participants feared that AI might process this information wrongly and consequently provide inaccurate recommendations. Eventually, the reliance of users on the false decisions provided by the AI may result in worsening the situation of the user.

## 4.3. Memory Avatars

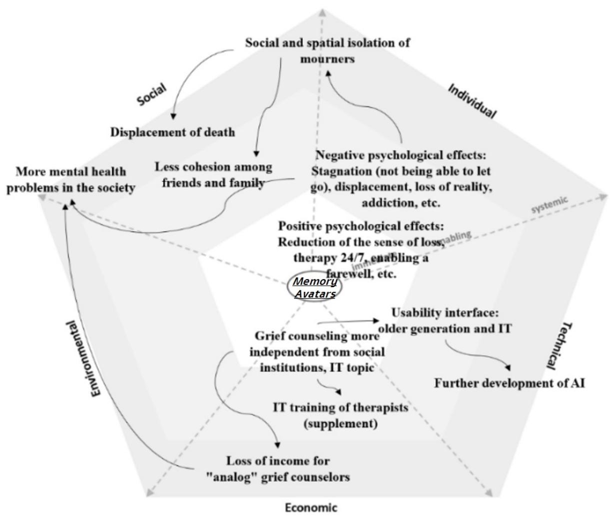


Figure 3. SusAD: Memory Avatars

The results from the workshop on Memory Avatars revealed worrying concern about the use of these technologies. For instance, although participants mentioned that the bereaved individuals might achieve temporal psychological satisfaction, such as minimizing the pain associated with losing a loved one through the use of memory avatars is without a series of negative effects. Thus, the impact of minimizing the pain in the death of a loved one leads to series of negative consequences. For example, this could lead to a person living in the past and not coming to terms with reality. As a result, individuals become dependant on technology as a coping mechanism which leads to severe problems such as mental health problems, addiction and abuse of technology.

As social animals, people gather to share happiness for a positive milestone in one's life or empathize when there is a tragedy. However, using Memory Avatars will decrease social affection, and people will stop caring about each other. Soon, death becomes an unspoken word the social fabric that holds families and friends together begins to fall apart. As such, this might even lead to worse social vices like kidnapping, unjustified killing without prosecution, and failure to demand justice when someone commits murder.

Economically, participants mentioned the loss of income as a negative impact. In other words, participants feared that AI Memory Avatars might replace therapists whose work is to ensure that people professionally deal with grief and sorrow. This follows a general skepticism that some people think in the future, AI will displace humans and render a lot of people unemployed. Even if health professionals proving therapy for the bereaved person decides to use memory avatars as a complement to their work, it will create economic inequality as not every person can afford to buy memory avatars.

From the technical perspective, although people of all ages can die at any time, death is witnessed often among older adults. Hence, the participants envisaged that IT companies would have to place their best bet on the elderly to establish a working system.It was also mentioned that this type of technology requires heavy investment in AI to facilitate the enhancement of Memory Avatars. Nonetheless, a technical problem or fault in the functioning of a Memory Avatar could lead to a far-reaching negative consequences like heart failure when the technology malfunctions and suddenly a person has to deal with the bitter reality.

## 4.4. Survey

Overall, 38 participants completed the survey questionnaire in all three workshops: Nine in Autonomous Driving, 5 in Music Composition and 24 in Memory Avatars. For ease of understanding, we classified the results into a ratio of one to ten.

**Change of attitude towards the subject:** Seven to eight out of ten participants stated that their perception of sustainability impacts had changed as a result of the workshop. This was evident as participants became aware of for example: "Wide range of possible impacts", "Sustainability is multidimensional (my focus was on ecology)" and "different dimensions in the framework used". In particular, participants became aware of the social and technical dimensions, which they had neglected before. Approximately nine out of ten participants gained insights they had not had before.

**Comprehension and benefit from the SusAF:** For about nine out of ten participants, the workshop was somewhat understandable and comprehensible or very understandable and comprehensible. All participants agree that the workshop should be repeated over time. The question of whether the value of the results is commensurate with the time spent was answered in the affirmative by eight out of ten of the participants. Seven to eight out of ten participants felt that their AI would benefit from sustainability integration, and nearly all participants would recommend the workshop to others. All participants answered the question, "Do you think that the use of the Extended SusAF will have an impact on products and services in terms of sustainability in your company or in other companies?" with a yes. Eight out of ten participants indicated that they would be interested in performing a similar analysis for future AI software.

**Future studies:** For further work on AI software, nine out of ten participants would like to see offers for education and training for employees in the area of sustainable AI as well as communication between science and industry. Eight out of ten participants would like to see (interactive) material on SusAF and other sustainability tools. Four out of ten of the participants would like to see public funding programs for companies tackling sustainable AI.

## 5. Discussion

We first answer our RQ with three lessons learned. Then, we explain the threats for validity of our study.

### 5.1. Lessons learned

**Lesson 1 - Align the knowledge:** In addition to an introductory presentation on the respective market environment of AI, all stakeholders must have a basic understanding of what AI is. Additional explanation time should be allowed for those participants who have not or hardly dealt with AI so far but still want to contribute to the discourse. A general understandable explanation should be built into the introductory presentation accordingly. In each workshop, in order to include different perceptions and to bring new views, the participants dealt with the impacts of AI on the field under investigation. A sufficient number of stakeholders from different areas is elementary to collect valuable results.

**Lesson 2 - A multidimensional tool first, focusing tool(s) second:** In each workshop, some participants had already dealt with a greater or lesser extent with individual impacts of AI on the field under investigation. The survey shows that a multidimensional sustainability analysis broadens the view of the impacts and thus expands them. Dimensions were taken into account that they had not been considered before. As an example, Autonomous Driving has an impact on social interaction (fewer encounters due to the elimination of public transport). When the impacts of Autonomous Driving are discussed, the focus is almost exclusively on the economic and ecological dimensions (also in scientific studies). As soon as the overview has been expanded with the help of tools such as the SusAF, it is possible in a second step to select tools that sharpen the view, e.g., about fairness. If the analysis were to focus on one dimension or selected aspects within a dimension right in the beginning, this would lead to disregard or, in other words, too much would fall by the wayside. Additional issues will need to be explored in the future, such as privacy concerns and third-party impacts, especially for Autonomous Driving.

**Lesson 3 - Create incentives for sustainable enterprises:** This lesson is addressed to policymakers in the AI sector but can probably be applied to almost all industries. Incentives should be created so that companies strive for sustainability in their products and services. Science should not ignore the fact that the industry must think and act economically to be able to exist. The questionnaire picks up on this fact and makes it clear that there is interest on the part of the industry in support, e.g., in the form of further training for employees, provision of prepared materials, the establishment of funding programs, and other financial reliefs.

### 5.2. Threats to validity

The conducted study is an explorative mixed-method empirical research. However, the focus is on the explorative and qualitative parts. Therefore, we do not intend to achieve generalisability but to generate answers to our RQ. Our findings are the first step and should be useful for follow-up studies that contribute to verification and deepening. Nevertheless, we have focused on collecting a comprehensive data set by selecting participants of different gender, age, and academic and professional backgrounds. Regarding the risk of reliability, the authors analyzed the workshop results separately. When discrepancies arose, we discussed them until we reached a consensus.

A risk to construct validity is that the workshop participants did not understand their tasks properly. Therefore, we used written and verbal instructions. Additionally, we allowed workshop participants to ask questions at any time. Moreover, we used an already empirically evaluated tool (SusAF).

The threat that the workshop participants perform their tasks is socially desirable, especially as a reactive bias to the presence of the researcher and the other participants cannot be argued. To reduce this threat, we assured the participants of their anonymity when dealing with the data. Participants were also free to use their cameras and any names they chose during the workshops. Nonetheless, this is a threat challenging to avoid when working with groups of participants.

Confounding factors cannot be ruled out. One factor affecting the results of the workshops is the differences in knowledge regarding AI. Although, we aim to ensure a similar perspective on AI and knowledge of sustainability and the SusAF method by delivering introductory sessions and instructions to the workshop participants and by selecting participants with at least a basic knowledge regarding AI.

During these Covid-driven times, we cannot exclude the risk of confounding factors caused by the workshops that took place online. If the workshops are to be repeated, we recommend that a comparison will be made where all participants are physically present. A major confounding factor and threat to validity are the participants themselves. A different set of workshop participants might lead to different outcomes. To minimalize this threat, we selected a diverse set of participants in age, experience, and basic knowledge regarding AI, and we included experts (e.g., musicians and researchers). All participants are European and have an academic background.

Similarly, a threat to internal validity is the biased selection of the participants for the workshops. As we selected people with diverse backgrounds and based on PD principles, we tried to minimize this threat. However, the participants are biased as we did not select them randomly based on the population. Regarding the participants themselves, we want to clarify again that we do not want to generalize our results, and we do not compare the outcome. Instead, we want to show the feasibility of our approach and provide initial steps toward sustainable AI.

Our study made it clear that using the SusAF in the participatory workshops was effective in creating sustainability awareness in the AI-related topics. Future studies could extend the use of this framework in other AI related activities.

## 6. Summary

Software companies should be aware that they create powerful tools that have a profound and multidimensional impacts. For this reason, they should be self-critical of their decisions and aware of their responsibility to minimize the risk of unintended negative impacts. Our three workshops have shown partly unexpected positive and negative short and long-term impacts in the different dimensions that companies need to address. Key areas of focus to guide the initial movement toward sustainable AI products and services are, for example, that Autonomous Driving which could lead to more social isolation as public transportation loses market share. AI in Music Composition could affect a user's mental health status. Finally, that Memory Avatars could extend the grieving process. The SusAF is thus particularly suitable as an instrument for assessing the possible impacts of a product, service or system. Subsequently, tools that deepen individual aspects can be selected and applied. Software companies are likely to show more interest in sustainability issues in future as many stakeholders become more aware about sustainability. It is essential to emphasize a heterogeneous and diverse composition of stakeholders based on PD to be able to cover all dimensions adequately. Just as non-technical stakeholders are expected to acquire a basic understanding of AI to participate in the discourse, requirements and software engineers must open up to new fields. On this basis, a profitable exchange can take place.

## References


Agarwal, A., Agarwal, H., & Agarwal, N. (2022). Fairness score and process standardization: Framework for fairness certification in artificial intelligence systems. *ArXiv*, *abs/2201.06952*.

Ahmad, K., Bano, M., Abdelrazek, M., Arora, C., & Grundy, J. (2021). What's up with requirements engineering for artificial intelligence systems? *2021 IEEE 29th International Requirements Engineering Conference (RE)*, 1–12. https://doi.org/10.1109/RE51729.2021.00008

Amershi, S., Begel, A., Bird, C., DeLine, R., Gall, H., Kamar, E., Nagappan, N., Nushi, B., & Zimmermann, T. (2019). Software engineering for machine learning: A case study. *Proceedings of the 41st International Conference on Software Engineering: Software Engineering in Practice*, 291–300. https://doi.org/10.1109/ICSE-SEIP.2019.00042

Angell, R., Johnson, B., Brun, Y., & Meliou, A. (2018). Themis: Automatically testing software for discrimination. *Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, 871–875. https://doi.org/10.1145/3236024.3264590

Becker, C., Betz, S., Chitchyan, R., Duboc, L., Easterbrook, S. M., Penzenstadler, B., Seyff, N., & Venters, C. C. (2016). Requirements: The key to sustainability. *IEEE Software*, *33*, 1–1. https://doi.org/10.1109/MS.2015.158

Berendt, B. (2019). *Paladyn, Journal of Behavioral Robotics*, *10*(1), 44–65. https://doi.org/doi:10.1515/pjbr-2019-0004

Bogina, V., Hartman, A., Kuflik, T., & Shulnet-Tal, A. (2021). Educating software and ai stakeholders about algorithmic fairness, accountability, transparency and ethics. *International Journal of Artificial Intelligence in Education*, 1–26. https://doi.org/10.1007/s40593-021-00248-0

Brun, Y., & Meliou, A. (2018). Software fairness. *Proceedings of the 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, 754–759. https://doi.org/10.1145/3236024.3264838

Chouldechova, A., & Roth, A. (2020). A snapshot of the frontiers of fairness in machine learning. *Commun. ACM*, *63*(5), 82–89. https://doi.org/10.1145/3376898

Duboc, L., Betz, S., Penzenstadler, B., Akinli Kocak, S., Chitchyan, R., Leifler, O., Porras, J., Seyff, N., & Venters, C. C. (2019). Do we really know what we are building? raising awareness of potential sustainability effects of software systems in requirements engineering. *2019 IEEE 27th International Requirements Engineering Conference (RE)*, 6–16. https://doi.org/10.1109/RE.2019.00013

Fagbola, T. M., & Thakur, S. C. (2019). Towards the development of artificial intelligence-based systems: Human-centered functional requirements and open problems. *ICIIBMS 2019*, 200–204. https://doi.org/10.1109/ICIIBMS46890.2019.8991505

Galaz, V., Centeno, M. A., Callahan, P. W., Causevic, A., Patterson, T., Brass, I., Baum, S., Farber, D., Fischer, J., Garcia, D., & et al. (2021). Artificial intelligence, systemic risks, and sustainability. *Technology in Society*, *67*, 101741. https://doi.org/https://doi.org/10.1016/j.techsoc.2021.101741

Gillespie, N., Lockey, S., & Curtis, C. (2021). *Trust in artificial intelligence: A five country study*. The University of Queensland; KPMG. https://doi.org/10.14264/e34bfa3

Grunwald, A. (2020). *Technikfolgenabschätzung – eine einführung (technology assessment - an introduction)*. edition sigma. https://doi.org/10.1017/9781108869577

Khakurel, J., Penzenstadler, B., Porras, J., Knutas, A., & Zhang, W. (2018). The rise of artificial intelligence under the lens of sustainability. *Technologies*, *6*(4). https://doi.org/10.3390/technologies6040100

Kostova, B., Gurses, S., & Wegmann, A. (2020). On the interplay between requirements, engineering, and artificial intelligence. In M. Sabetzadeh, A. Vogelsang, S. Abualhaija, M. Borg, F. Dalpiaz, M. Daneva, N. Condori-Fernández, X. Franch, D. Fucci, V. Gervasi, & et al. (Eds.), *Joint proceedings of refsq-2020*. CEUR-WS.org.

Penzenstadler, B., Duboc, L., Akinli Kocak, S., Becker, C., Betz, S., Chitchyan, R., Easterbrook, S., Leifler, O., Porras, J., Seyff, N., & Venters, C. C. (2020). The susaf workshop – improving sustainability awareness to inform future business process and systems design, 13. https://zenodo.org/record/3676514#.YkFxzC8RpT7

Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., Breazeal, C., Crandall, J. W., Christakis, N. A., Couzin, I. D., Jackson, M., & et al. (2019). Machine behaviour. *Nature*, *568*(7753), 477–486.

Rainie, L., Anderson, J., & Vogels, E. A. (2021). *Experts doubt ethical ai design will be broadly adopted as the norm in the next decade*. Pew Research Center.

Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J.-F., & Dennison, D. (2015). Hidden technical debt in machine learning systems. *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2*, 2503–2511.

Sharma, S., Henderson, J., & Ghosh, J. (2019). Certifai: Counterfactual explanations for robustness, transparency, interpretability, and fairness of artificial intelligence models. *ArXiv*, *abs/1905.07857*.

Simonsen, J., & Robertson, T. (2012). *Routledge international handbook of participatory design*. Routledge.

Suárez, J., & Varona, D. (2021). *The ethical skills we are not teaching: An evaluation of university level courses on artificial intelligence, ethics, and society*. University of the Western Cape.

Vöneky, S. (2020). Key elements of responsible artificial intelligence: Disruptive technologies, dynamic law (M. Löwisch, T. Würtenberger, & C. Feldmann, Eds.). *Ordnung der Wissenschaft*, (1), 9–22. https://doi.org/10.17176/20200103-154137-0

Wynsberghe, A. (2021). Sustainable ai - ai for sustainability and the sustainability of ai. *AI and Ethics*, *1*. https://doi.org/10.1007/s43681-021-00043-6