Sustained Enablement of AI Ethics in Industry

Martina Flatscher(✉), Anja Fessler, and Isabel Janez

ZF Friedrichshafen AG, 88038 Friedrichshafen, Germany
martina.flatscher@zf.com

Abstract. Artificial Intelligence (AI) has become an increasingly pervasive technology in various industries, offering numerous benefits such as increased efficiency, productivity, and innovation. However, the ethical implications of AI adoption in industry have raised concerns and AI ethics has emerged as a critical field of study, focusing on the trustworthy development, deployment, and use of AI technologies. In this paper, we explore an AI Ethics concept with a particular focus on sustained enabling factors to guide organizations in navigating the ethical challenges associated with AI adoption.

Keywords: Artificial Intelligence · Ethics · Trustworthy AI · Innovation · dynamic Framework · Enablement · Measurement

1 Introduction

The rapid advancement of AI has led to its widespread adoption across various industries, including healthcare, finance, manufacturing, transportation, and more. AI technologies, such as machine learning, natural language processing, and computer vision, have enabled organizations to automate processes, gain insights from data, and enhance decision-making. Foundation models like Chat GPT, as one of many applications and a very prominent example shows how much AI has arrived in society and industry. Along with the benefits, the increasing use of AI has raised ethical concerns related to its impact on society, economy, and individuals. Bender et al. [1] illustrates the importance of the consideration of AI Ethics because of the risks involved with AI Applications like large language models. In terms of use and the scope of AI Applications there are various level of criticalities, related, and to the potential to harm people through the usage of personal data or direct involving of people using the AI Applications [2], which need to be addressed accordingly [3, 4].

However, there are no clear standards for the appropriate handling of ethical considerations in industry. Position papers reflect that current movements target to become more and more ethical, as industries are highly requiring ethical procedures [5–7]. New terms are evolving around AI Ethics like Trustworthy AI and Responsible AI [8]. The industry is trying come up with ethical frameworks, standard processes, and best practices that align with the values of organizations and society to deal with new challenges coming with the dynamics of AI [9–11]. However, industry struggles in holistically manage AI Ethics. This paper proposes an approach to frame AI Ethics and enable it in a feasible way.

M. Yilmaz et al. (Eds.): EuroSPI 2023, CCIS 1890, pp. 3–19, 2023.
https://doi.org/10.1007/978-3-031-42307-9_1

2 State of the Art

To understand what AI Ethics compromises it is of importance to first clarify and narrow down related terms that appear in the research, and based on this, to expand the relevant subject areas.

Section 2.1 gives an overview of used terms related to AI Ethics and comes up with a definition of terms to build a common understanding. Section 2.2 and 2.3 grasps the scope of the study area in this paper summarizing AI Ethics values and enablement approaches.

2.1 Terms in Research Around AI Ethics

In industry the term AI Ethics isn't properly defined, therefore related terms will be clarified in the following:

The term *Trustworthy AI* (TAI) is often used to ensure that AI systems are moral, open, and responsible. Accuracy, robustness, openness, explainability, human management and monitoring, justice, the elimination of bias, and security are requirements for TAI systems [2–7]. It turns out that TAI is more likely used when it comes to values that engineering and governance departments can fulfill and work on by developing appropriate methods and processes.

Responsible AI, inter alia used by Gartner [2], is used when the human centric approach is of major importance [3]. Beyond algorithmic fairness, responsible AI algorithms cover significant facets of AI that may contribute to avoid AI's apathetic behavior, such as societal effect and human rights protection [4]. Concerns exist over AI's capacity for moral behavior and decision-making, and several ethical frameworks, guidelines, and rules have lately been implemented to address these concerns [5].

The term *AI Ethics* includes more than just the engineering point of view, referring to standards, guidelines, and frameworks that direct the creation and responsible usage of AI systems. An operationalization of AI Ethics will consider and incorporate all components [6].

In the following we will use the term TAI when it comes to the definition of values, which ensure ethical behavior in AI development. Thereby, we are referring to the HLEG Group and its definition, shown in Sect. 2.2. In distinction to this, we use the term AI Ethics in the further course for the holistic view on the topic. In this context, we will examine how industry can be guided towards a more ethical behavior of AI development, deployment, and use.

2.2 Trustworthy AI Values in Research

Due to wide adoption of AI, society and industry require ethics in AI. Many principles and guidelines for ethical AI have been issued by private companies (e.g. 2019:84 officially published guidelines for AI) [7–11], research institutions [12] and public sector [13]. What they all have in common is that they span the complex field of AI Ethics into values [14].

Many authors come up with a bunch of AI Ethic values [12, 14–20], which the *European Group on Ethics in Science and New Technologies* summarizes the most

comprehensively. The *Trustworthy AI* values of the HLEG Group are the following seven values shown in Fig. 1 below [2, 12]:

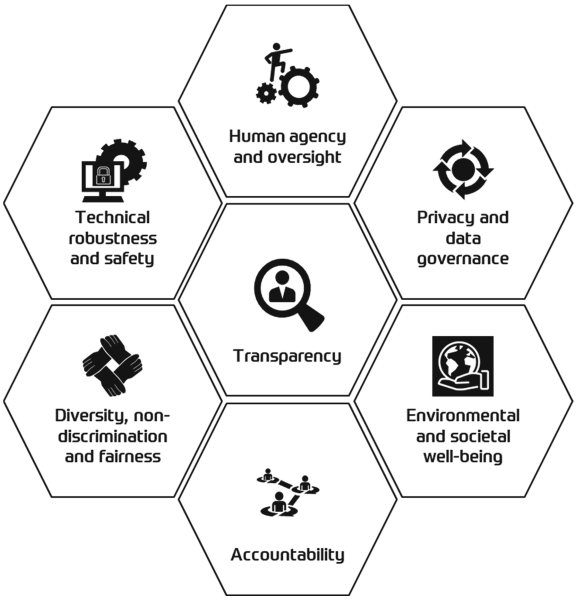

Fig. 1. TAI values according to HLEG [21]

TAI values can be differentiated into different levels of tangibility. For a better understanding of the values and their themes, Hagendorff distinguishes between quantifiable values (explainability, privacy and fairness) and unseizable values, due to the topic's complexity. Farther he points out that the scope of AI Ethics is not adequately grasped by often idealized, quantifiable, and calculable forms of quantifiable values and needs a holistic view to ensure AI Ethics [15].

TAI requires implementation on a technical and governmental level. In research much effort is devoted to investigating more and more approaches and tools to implement *quantifiable values* [22]. However, governmental aspects, which fall under the *unseizable values* are far more difficult to grasp and will be investigated in an enabling context in the next section.

2.3 Enablement Approaches in AI Ethics

Frequently research, industry, and government came up with enablement approaches for AI Ethics. Identified approaches can be categorized into:

Expand AI Ethics into *Action Fields* to make complexity tangible: National strategy documents derive *Action Fields* to grasp AI Ethics in a governmental context. Amongst many nations, Europe, the USA and China [14, 23, 24] are major player in setting

standards. HLEG provides a framework for outlining a set of self-assessment questions specifically regarding the *TAI values* mentioned in Sect. 2.2 [25].

From industrial perspective KPMG has developed the AI in Control Framework to help organizations build greater confidence and transparency along the AI lifecycle - from strategy to development - through tested AI governance constructs as well as methods and tools. It also provides some of the key recommendations and leading practices for implementing AI governance, conducting AI assessments, and developing continuous AI monitoring and visualization [26].

Risk based approach: The European Commission, as the first major regulator that proposes a law on AI, assigns AI applications to risk categories. As an attempt to advance digitization in the EU and make it competitive in an international comparison the AI Act recommends dedicated actions regarding their risk class [27]. At a supranational level another regulatory initiative comes from UNECE, to use AI in trade facilitation [13].

AI Committee: Setting up AI Committees to centrally coordinate AI Ethics initiatives like AI Use Case Portfolio, AI opportunities, AI Risk Management, etc. [18, 23].

Standards: Using Standards, where AI Ethics is ensured by step-by-step approach [28] IEEE presents the first industry standard "IEEE 7000-2021", that addresses AI Ethics operationalization [29].

AI Ethics Measurement approaches: Enabling AI depends beyond a feasible approach heavily on knowing your current status and evaluating it. This should incorporate regular assessments and evaluations to ensure that TAI is not only implemented but also maintained and continuously improved. These assessments should be done in a transparent and accountable manner to build trust and confidence in the TAI system [25]. Defining specific evaluation criteria helps to reduce complexity and to get management attention [22, 30, 31].

The AIC4 allows an independent auditor to conduct an attestation engagement on the AI service's compliance with the criteria [32]. The Bertelsmann Stiftung has laid the foundation for potential measurability in the use of AI through its ethical label approach. AI companies can use such a label to publicly communicate the quality of their products. The label improves the comparability of products on the market for consumers and AI-using organizations and provides a quick overview of whether an algorithmic system meets the required ethical standards. The principles of the HLEG Group are adopted and endorsed [33]. ETAMI also uses AI Auditing and Ethics Label to align trustworthy and ethical AI between academia, industry, and society [34].

Although the engineering side has been partially worked out, there is a lack of a holistic solution and a lack of a strategic approach [17, 35, 36], which is why the industry has not yet fully integrated AI Ethics.

3 Research Question and Methodology

The literature review showed that *TAI values* and enabling approaches evolve, but industry is struggling to actively ensure AI Ethics holistically.

The research question that emerges from this review can be formulated as follows:

How to live a successful approach to enable AI Ethics in industry?

To answer this question, we propose a concept for AI Ethics and recommend a sustained enablement approach and a measurement concept. This concept is applied to industry that is developing AI and applying it in systems.

Section 2.1 is based on a literature review on AI Ethics beginning with terms definitions and in-depth literature review on TAI (Sect. 2.2) and enablement approaches (Sect. 2.3). The advancements of ensuring AI Ethics by technical solution in coding has been well-explained, but the incorporation of governmental procedures has not yet been well addressed in the literature. Therefore, this paper transfers findings from literature into an AI Ethics approach enhanced through feasibility aspects considering the industrial context. Section 4.1 introduces the necessary adaptation of theory for industry. The AI Ethics Concept is divided into *TAI values* (Sect. 4.2) and recommended AI Ethics action field (Sect. 4.3). Section 4.4 puts it all in an industrial context highlighting the existing correlations within an organization. Chapter 5 deals with the feasibility of AI Ethics, where Sect. 5.1. Elaborates on a living integrated process approach and Sect. 5.2. Comes up with a measurement approach. This concept is about to be applied to industry that is developing AI and applying it in systems.

4 Development of the Enablement of AI Ethics

This work aims to define a concept for bringing AI Ethics into industry. To succeed in such a practical process Sect. 4.1 elaborates on the introduction which various aspects must be considered in an AI Ethics concept to guarantee holism. In the following three Sects. 4.2–4.4 the elements and success factors for an AI Ethics concept are presented.

4.1 Introduction to a Practical AI Ethics Concept

Literature findings show a variety of approaches and tools to handle the challenges of AI Ethics. However everyday practice teaches us, that theoretical constructs like technical standards, best practices, and ethical principles for AI, often have to be adapted to industry in order to be successful and feasible. This is because researchers frequently have a thorough knowledge of AI's technical and ethical issues but may have little practical expertise with commercial operations and real-world applications. Therefore, findings from research needs to be transferred into very concrete action plans useable in industry.

These action plans must in turn be individually adapted to the specific practical considerations. Also, in terms of business operations, the diversity of AI application fields must be considered.

Within industry various fields of operating levels need to be considered regarding AI Ethics (further elaboration in Sect. 4.4 see Fig. 4). Engineering teams on the one hand ensure that the development of code and algorithms are transparent, fair, and explainable (see *quantifiable values* Sect. 2.2), to not reinforce or amplify biases. This requires a deep understanding of how AI algorithms work and the ability to identify and address biases and errors. Technical teams also need to consider how to manage data privacy and security, and how to mitigate the risk of data breaches or other security threats.

At the same time, governance teams need to ensure that their regulations and policies address the ethical considerations of AI, including issues such as privacy, security, and accountability (see *unseizable values* Sect. 2.2). This also requires the ability to identify and address ethical risks and concerns.

Additionally, another challenge of applying TAI in industry is the need to balance the competing demands of innovation and ethics. While TAI is critical for ensuring that AI benefits society, it can also place significant constraints on the development and deployment of AI systems. This can create tensions between the technical and ethical considerations of AI and the practical considerations of business operations.

4.2 Trustworthy AI Values

AI Ethics needs to be based on a common understanding regarding what values stand behind it. As described in Sect. 4.1, enabling via a theoretical construct only succeeds if it is made tangible for industry. Consequently, [10, 33] defined *AI Ethics Guidelines* that break down the *TAI values* increase employee understanding in clarifying the *TAI values* [10, 33]. Hence the identified *TAI values* [21] in Chapter 2 are clarified in the following, by approaching it by listing further subject areas (named *principles within values* in Fig. 2), which fall under the respective value. The interpretation of these values can vary depending on the application context, the cultural region and the stakeholders involved. Example key tasks are listed in the bottom column of the table to derive *TAI values* on a work level. Tasks have technical and governmental characteristics. Figure 2 illustrates the subject areas of how TAI can be grasped in industry. This list of principles and tasks is a glimpse into the industry, as the list does not claim to be complete.

TAI Values	Human agency and oversight	Technical robustness and safety	Privacy and data governance	Transparency	Diversity, non-discrimination, fairness	Environmental and societal well-being	Accountability
Principles within TAI Values	- Fundamental rights compliance - Human review - Human control of technology - Human review of automated decision	- Security by design - Predictability - Resilience to attack - Cybersecurity - Risk reduction - Reliability - Transparency - Standardization	- Control over use of data - Right to rectification - Right to erasure - Privacy by design - Access to data	- Explainability - Interpretability - Traceability - Open-source data and algorithms - Open government procurement - Right to information	- Fairness - Equality - Inclusiveness in impact - Non-discrimination and the prevention of bias - Inclusiveness in design - Representative data (ODD) - High-quality data	- Environmental friendliness, protection and well-being - Resource saving - Human rights - Sustainability - Social impact	- Verifiability and replicability - Impact assessment - Liability and legal responsibility - Evaluation and audit requirement - Ability to appeal - Remedy for automated decision - Monitoring
Related tasks	- Ability to opt out of automated decision - Ensure human control in critical AI-decision use cases	- Integrate stakeholder along the whole AI life-cycle to develop explainable, safe and robust - Actively shaping the landscape of norms related to safe and secure AI	- Ability to restrict processing - Comply with privacy and data governance policies	- Notification when interacting with an AI - Notification when AI makes an individual-related decision - Regular reporting	- Build fair AI applications - Check for unbiased data - Monitor usage of AI in human-critical use cases → AI use case portfolio	- Monitor and consider usage of AI for environmental and societal well-being → AI use case portfolio	- Risk management - Assignment of responsibility - Recommendation for new regulations - Establishment of a supervisory authority - Governance & life-cycle mgmt. principles for documenting, developing, verifying, monitoring, incidence handling and auditing for AI

Fig. 2. Exemplary TAI Value principles and related tasks oriented on [10, 37]

Ultimately, the successful application of TAI in industry will depend on the careful consideration of these foundational values and their interpretation in different application fields. This will require ongoing dialogue and collaboration between stakeholders from industry, academia, and civil society.

4.3 AI Ethics Action Fields

Despite the importance of engineering skills that ensure AI Ethics on a technical level, the enabling of purely technical AI Ethics aspects is not further elaborated in this paper, as this is already well covered in literature. This paper addresses the missing concrete approaches for governmental tasks and comes up with AI Ethics *Action fields*, inspired by the research study in Chapter 2. These seven *Action Fields* represent foundation pillars to holistically work on. Each action field can be spanned into several subcategories illustrated in Fig. 4. For a better understanding of how to scope the *Action Fields*, only the most relevant subcategories are mentioned in the Section that follows [38]. Please note that the listed subcategories do not claim to be exhaustive but should rather be seen as inspiration.

AI Literacy/Education: Practice shows that AI often raises ethical concerns due to a lack of knowledge in the company. Employees are simply often not aware of what an ethical AI model requires or what implications AI applications can have if used incorrectly. It is therefore of enormous importance to train the company specifically and to point out AI Ethics aspects or even guided through AI Ethics training courses. This involves providing training and education to employees and stakeholders on the technical and ethical considerations of AI systems. This could include providing training on data privacy and cybersecurity, as well as education on ethical principles and values. Training makes aware of *TAI values* in the company and thus they specifically can be considered through upskilling in development. Through training and associated upskilling, employees are made aware of the *TAI values* in the company and can specifically take it into account in their development. Practical knowledge management is one subcategory that contributes to the important dissemination of state of knowledge around AI Ethics. Furthermore, it's building trust as it creates a relationship constructs and commitment to the topic.

AI Risk Management/Mitigation: Being aware of the risks of using AI sharpens the handling of the same technology. Here it is primarily a question of avoiding humaninjuring applications beforehand, but also of uncovering competitive opportunities through not using AI. Furthermore, the AI Lifecycle in its total has to be covered in a continuous process where unforeseen harm must be thoroughly assessed and managed. Risk mitigation on the one hand uses organizational tools like communication and awareness campaigns to avoid a misuse of AI through lacking knowledge, on the other hand the improvement of an AI model to become more robust and transparent also contributes to reduce AI risks on an engineering level.

AI Awareness/Communication: The more employees understand how an AI model works, the more engineers (AI Hub) understand their toolset of improving their AI use cases regarding AI Ethic concerns, the more ethical a company acts in using AI. This action field is clearly linked to AI Literacy/Education as this contributes to increasing AI Awareness. Nice side but decisive effect is, that awareness creates trust, which is crucial for AI users to accept the technology, for engineers to consider AI in product/process development [41] and to realize the full potential of AI in the long term [20]. Psychological trust models name performance, reliability, prediction and explainability as decisive factors in building trust for AI applications [42, 43]. Especially transparency can be called an important factor for trust in AI [44].

AI Compliance: The increasing number of AI regulation papers, norms, standards and laws requires a proper use of AI. Therefore, a compliance and governance approach is needed and needs to be rolled out across the group.

AI Policy: As mentioned in chapter 2, industry is coming up with a lot of AI Ethics position papers [14]. Outlining a compromise between too flexible and too strict Raab [47] argues that when prescriptive guidelines are implemented in a top down and nonflexible fashion, this gives the misleading impression that it is possible to take a formulaic approach to the application of ethical norms, principles and general rules to specific instances [47]. Therefore, AI Policy has on the one hand to build the rules and guidelines for the use of AI, on the other hand structure the monitoring to really fulfill AI Ethics requirements in the highly dependent corporate organization processes.

AI Auditing: As it is also intended in the *AI Act* an AI Auditing process has to centrally manage and monitor all AI activities throughout the company to fulfill *TAI Values* and thus ensure AI Ethics. The use of technology must be a comprehensible decision that is made based on risk management, also regarding the due diligence expected by law. Oversight mechanisms have to ensure that AI systems are developed and deployed in a responsible and trustworthy manner. This could include establishing an AI Ethics Committee to review, conducting regular audits and evaluations, and creating mechanisms for reporting and addressing ethical concerns [48, 49].

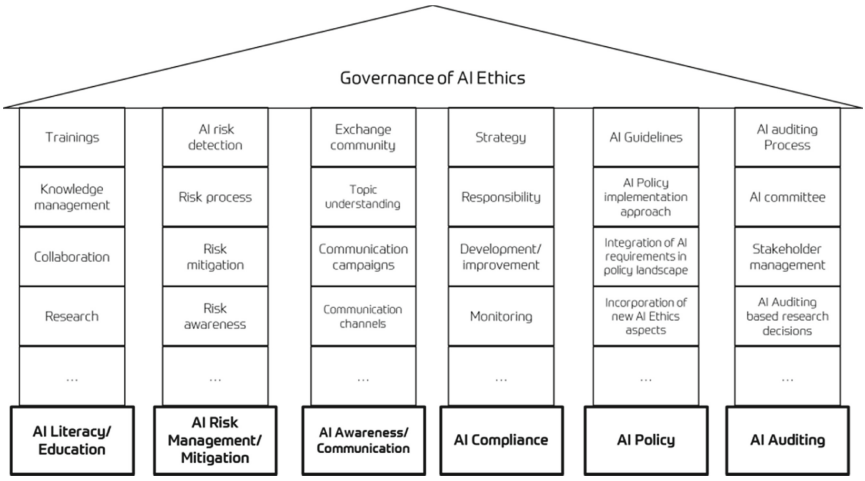

Fig. 3. Action Fields of AI Ethics

4.4 Embedding AI Ethics into Industrial Context

In industry AI is usually organized embedded in an interdependent corporate context. As depicted in Fig. 4 AI Ethics is composed of multidimensional aspects in a complex dependent inter-organizational relationship. The corporate top-down framework determines the strategic course of AI. At a technical level AI is developed and applied (upper

3 levels in the top box), whereas at a non-technical level AI has to be organized regarding Education, Governance, Communication, etc. (lowest level in the top box). To ensure *TAI values* it is essential to have a strong AI hub, where technical standards and best practices are elaborated and defined for developing and deploying AI systems in a way that is safe, reliable, and secure. This could include guidelines for data collection and processing, testing and validation, and cybersecurity. In addition, AI Ethics Guidelines are based on the *TAI values* to establish a respective value understanding.

The lower box illustrates the key elements including presented *Action Fields* from Sect. 4.3 for AI Ethics. Within the boxes, but also between the upper and lower box, there are strong dependencies and connections.

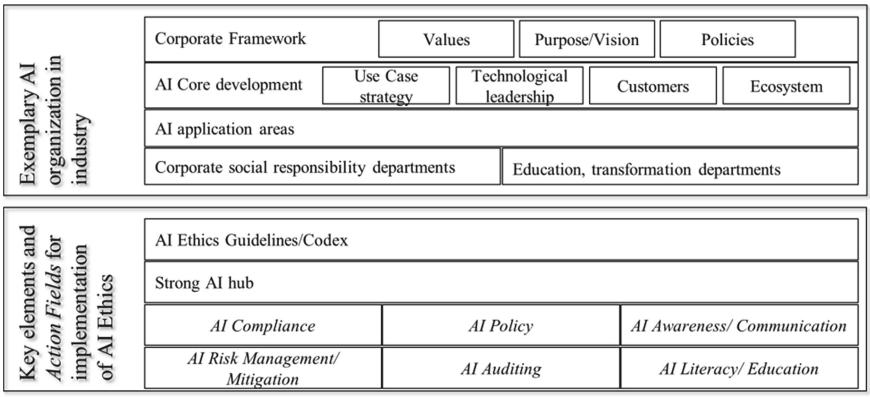

Fig. 4. Multidimensional nature of AI Ethics in an industrial context

The correlation between the *Action Fields* were mentioned in Sect. 4.3. The other given correlations will be explained in the following, as their consideration in an AI ethics concept is a crucial factor.

It is crucial to determine AI values strongly related to the corporate values and considering all stakeholder perspectives in order to align with the needs and values of stakeholders and to enable incorporation into the design and implementation of AI systems. Concrete *Action Fields*, shown in italics, are highly dependent on the organizational structure related departments are arranged. It is important to work interdisciplinary on the *Action Fields* breaking it down into subcategories (see Sect. 4.3) and adapt them company specific. This should involve collaboration between AI developers, industry experts, policymakers, and other stakeholders, including end-users and affected communities.

5 Enabling of AI Ethics

Lockey states that AI Ethics has a high similarity to innovation, thus the operationalization of ethical aspects requires a feasible concept that guides industry to success. This means that a one-size-fits-all approach to AI Ethics is not appropriate.

In the following Sect. 5.1 feasibility recommendations are given to succeed in AI Ethics. Based on that Sect. 5.2 describes a measurement approach to evaluate the progress.

5.1 Feasibility Aspects of AI Ethics in Industry

Given the fact that AI Ethics comes with intrinsic complexity, combined with high dynamics regards to the fast adaptation of AI Applications, a flexible operationalization is required with adaptation options in the sense of a living integrated process. As pictured in Fig. 5 road mapping, as a well-established strategic management tool [49–51], seems to be a good analogy for organizations to schematically show that companies have to walk through their *AI Ethics path* individually to become better.

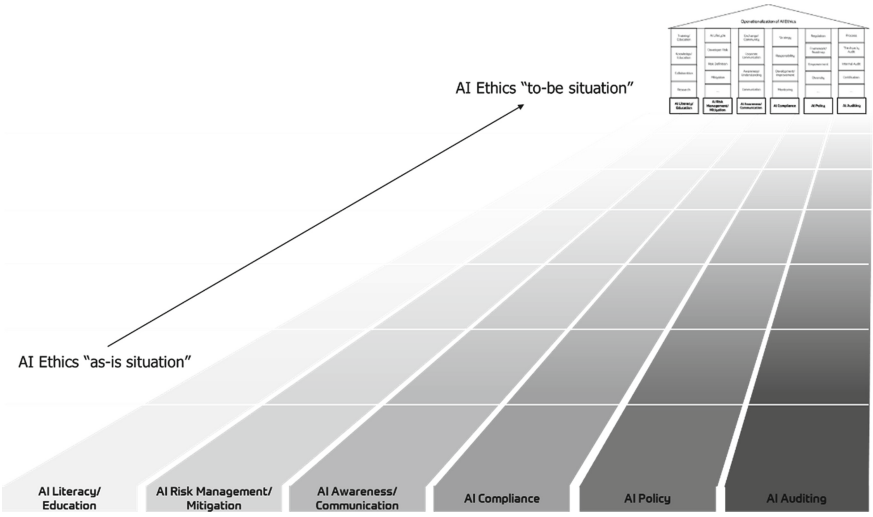

Fig. 5. Living integrated process to AI Ethics governance

In the following feasibility aspects are indicated, which show how the path to an AI ethics-compliant company can be successfully followed.

Clarity about relevant AI Ethics stakeholders and fruitful working mode: This aspect probably stands out the most, as AI Ethics, with its complex and multidisciplinary particularity, only subsist in companies, if the various stakeholders are identified. Companies must define stakeholder according to the various perspectives on the topic for example AI designers and AI developers of the AI system, Data Scientists, procurement officers or specialists, front-end staff, that will use or work with the AI system, legal/compliance officers, management, etc. [25]. The dynamic inter-company environment regarding employee turnover and priority shifts requires a consistent integration of these stakeholders in order to represent their interest, values and needs in the governmental (*Action Fields*) and technical procedures of AI Ethics within the AI *Action Fields*. Regarding

AI Governance tasks a dedicated AI Ethics Stakeholder Committee can facilitate the coordination of the AI *Action Fields.*

A factor and AI initiatives is meetings to work constructively and consistently on the action. The consistent work on the *Action Fields* increases the level of AI Ethics.

Stakeholder integrated definition of *TAI Values* and AI Ethics *Action Fields*: As pointed out in Chapter 4 the specific and stakeholder involving elaboration of *TAI values* as well as AI Ethics *Action Fields* depending on the company context and the organizational characteristic of AI Ethics, is of major importance. The recommended values in this paper can be used and adjusted specifically in order to get a holistic view on AI Ethics. This is achieved by uncovering identifying and tailoring concrete initiatives within the *Action Fields* that leads to fully ethical companies with regards to AI development [38].

Flexible operationalization and a living integrated process: Millar et al. [51] recognize that the leadership of disruptive innovation, as an intrinsic characteristic of AI Ethics, comes with volatility, uncertainty, complexity and ambiguity (VUCA) [52]. Companies have to deal with their own individual and personalized VUCA world, which requires tact and sensitivity [53, 54]. Therefore, undefined influences of the VUCA environment must be detected [55]. There are inter-company dynamics like shift of key persons or priorities. From an external perspective new trends and AI advancements needs to be managed because they involve risks and opportunities. Additionally, the AI specific challenges increase the complexity of the topic [56].

5.2 Measurement Approach for AI Ethics

The enablement of AI Ethics requires knowledge of the implementation status of the AI *Action Fields.* However, since the topic AI Ethics is very complex, the assessment must be made on a more concrete level than the *Action Fields.* As mentioned in Fig. 3 *Action Fields* must be defined in tangible subcategories. For example, the action field AI Education can be defined in AI training, AI Awareness and Collaboration and Research. These terms describe the Action Field in a holistic way and allow the topic to be examined from several perspectives. Subsequently, specific actions can be assigned to subcategories to get measurable elements. Using the subcategory AI Training as an example, contributing actions could be to establish an AI training strategy and provide AI trainings to employees.

Finally, indicators must be defined that make evidence of the activity within the *Action Fields.* In this example, the number of provided AI trainings is a quantifiable KPI for the AI Training action field. The indicators taken together provide a measurement standard on the basis of which a company can improve individually [53]. In this way stakeholders understand the objectives behind AI Ethics through the quantifiable indicators [56]. Figure 6 shows schematically how the measurement of AI Ethics is built.

Beyond the transparency of how a company performs in the defined *AI Ethics Action Fields* the consistent working on *TAI Values* itself has to be measured as well. Therefore, the measurement approach records them in the same way as the *Action Fields*. The *tasks* from Fig. 2 can be used to make the *TAI values* measurable. Indicators then need to be found to remain in the same evaluation logic as with the *Action Fields.*

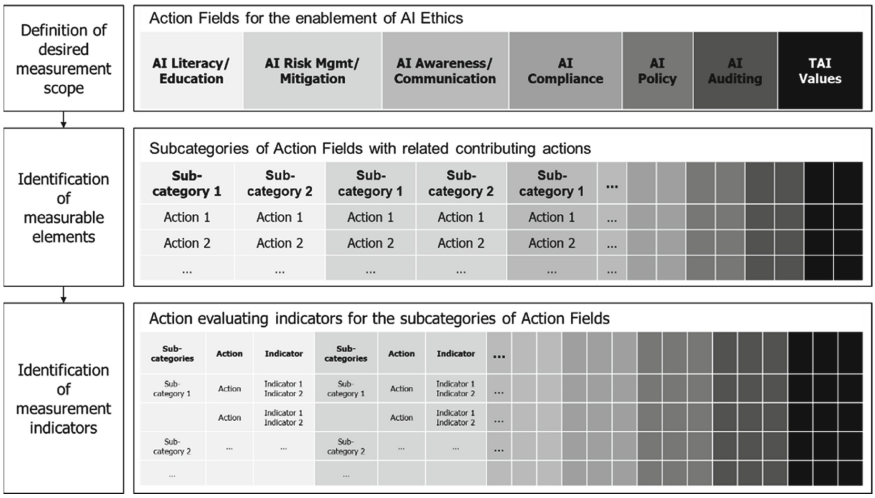

Fig. 6. Measurement Approach

Companies are not comparable with each other, as the scope of AI Ethics is framed and organized very differently. Therefore, companies must define individually their *Action Fields*, subcategories, actions, and indicators in order to get a realistic AI Ethics evaluation. In doing so, companies immediately see what they need to work on within the *Action Fields.*

6 Discussion

The motivation for defining a dynamic enablement concept for AI Ethics lies with the thought that societies will only ever be able to achieve the full potential of AI if trust can be established in its development, deployment, application and use [21]. If exemplarily, the general public doesn't trust autonomous cars, they will never replace common, manually steered cars [57]. That is why considerations of trust must begin in the manufacturing industry itself. Emaminejad N. [58], even defines it as an inevitable user acceptance requirement.

At the same time, it is of importance not to turn TAI into an intellectual land of plenty: it should not be perceived as an umbrella term for everything that would be nice to have regarding AI systems, both from a technical as well as an ethical perspective [53]. Since this can only go well if no worst-case scenarios arise.

To counteract this general societal issue the European Union is working on the *AI Act*. Given the need to address the societal, ethical, and regulatory challenges of AI, the EU's stated added value is to turn it into a competitive advantage under the banner of *Trustworthy AI*. This vision for AI-enabled technologies, which aims to mitigate potential harm and enable accountability and control, could set Europe apart from its global competitors. Also, it can also serve as a key component in strengthening the EU's digital sovereignty by giving European users more choice and control [54].

Though Castro [59], sees this critical and is wondering if Europe will Be Left Behind If It Focuses on Ethics and Not Keeping Pace in AI Development". He suggests "a delusion built on three fallacies: that there is a market for AI that is ethical-by-design, that other countries are not interested in AI ethics, and that Europeans have a competitive advantage in producing AI systems that are more ethical than those produced elsewhere".

In addition to the global and European perspective, the understanding across all industries, on the ethical level, should also be strengthened. This makes a cross-industry alignment regarding the interpretability of AI Ethics and the measurement of it, of interest. As a potential way to reconcile the complexity presented, as well as achieve a universal understanding on AI Ethics that is carried globally and across industries, we see the introduction of an AI Ethics label as an option. However, we see the challenges in introducing a uniform label that forms a global, Europe-wide, industry-wide consensus.

From an individual perspective, there needs to be further discussion on how to bring stakeholders and policy together with the technical perspective [59].

7 Conclusions

The wide adoption of AI makes it an indispensable part of the industrial context coming with complex ethical challenges that need to be addressed in a holistic approach. AI Ethics and its enablement needs to be tackled not only from an engineering point of view (strong AI hub) but also from a governmental side. To overcome the limitations of lacking standards and feasible approaches this work proposes a concept where key elements

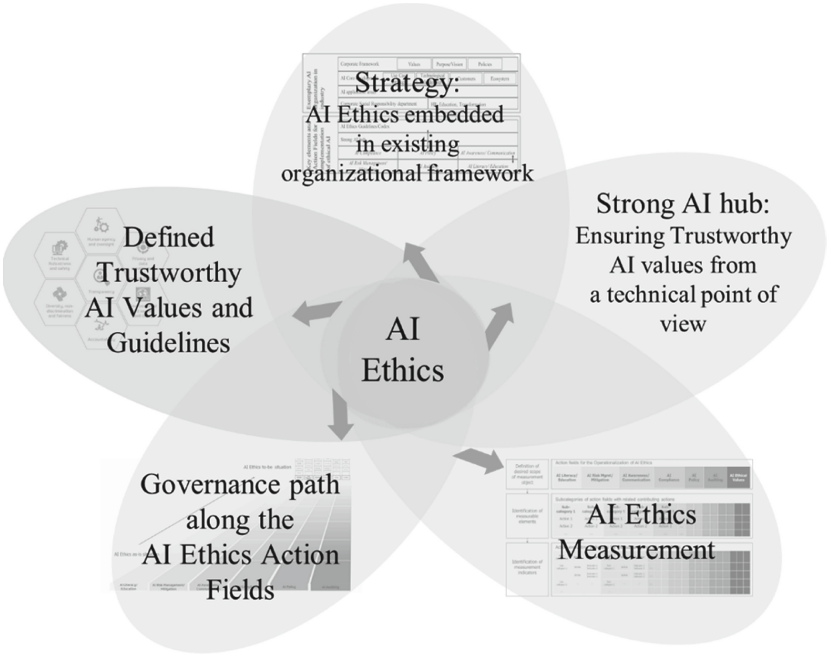

Fig. 7. Disciplines of an AI Ethics Enablement

related to AI Ethics, based on the research study, are framed a way, that companies can individually elaborate on their best AI Ethics performance. The framework is based on dedicated AI Ethics values that must align with and enhance the corporate values and needs to be considered in the whole corporate context. Framing AI Ethics *Action Fields* enable the operationalization to be dealt with holistically and tangibly.

Furthermore, practical implications are given to make the proposed concept as feasible as possible. The starting point to measure the topic of operationalization is an assessment of existing activities respectively related indicators that contribute to the *Action Fields*. Along the pathway companies assign more and more activities to the *Action Fields* and AI ethics becomes more and more complete. Figure 7 brings together the different disciplines that an AI Ethics Enablement requires to be feasible and successful.

8 Relationship with the SPI Manifesto

Process improvement and innovation where AI Ethics lies are at the very core of the long, successful *EuroAsiaSPI*^{2} history. The SPI manifesto created in this community defines the values and principles required to deploy SPI efficiently and effectively.

The enablement of AI Ethics needs to be integrated in processes, and continuously improved in order to adapt to the evolving understanding of organization maturity. The importance of transformation for organizations has clearly been evolving in the digital age, which is why this work is perfectly aligned with the SPI manifesto [60].

References

1. Bender, E.M., Gebru, T., McMillan-Major, A., Shmitchell, S.: On the dangers of stochastic parrots. In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, Virtual Event Canada, pp. 610–623 (2021)
2. Gartner: Gartner Identifies Four Trends Driving Near-Term Artificial Intelligence Innovation. https://www.gartner.com/en/newsroom/press-releases/2021-09-07-gartner-identifies-four-trends-driving-near-term-arti-ficial-intelligence-innovation. Accessed 19 Apr 2022
3. Mikalef, P., Conboy, K., Lundström, J.E., Popovič, A.: Thinking responsibly about responsible AI and 'the dark side' of AI. Eur. J. Inf. Syst. **31**(3), 257–268 (2022). https://doi.org/10.1080/0960085X.2022.2026621
4. Cheng, L., Varshney, K.R., Liu, H.: Socially responsible AI algorithms: issues, purposes, and challenges. J. Artif. Intell. Res. **71** (2021)
5. Lu, Q., Zhu, L., Xu, X., Whittle, J., Xing, Z.: Towards a roadmap on software engineering for responsible AI (2022)
6. Birhane, A., et al.: The forgotten margins of AI ethics. In: 2022 ACM Conference on Fairness, Accountability, and Transparency, Seoul Republic of Korea, pp. 948–958 (2022)
7. BMW Group: Seven principles for AI: BMW Group sets out code of ethics for the use of artificial intelligence (2020)
8. Google: AI at Google: our principles (2018)
9. Robert BOSCH GmbH: AI code of ethics: Bosch sets company guidelines for the use of artificial intelligence (2020)
10. Götz Manuel and Flatscher Martina: AI Ethics Guidelines. https://www.zf.com/ethical-ai. Accessed 27 Apr 2023

11. Microsoft: Responsible AI principles from Microsoft. https://www.microsoft.com/en-us/ai/responsible-ai?activetab=pivot1:primaryr6. Accessed 27 Apr 2023
12. Hagendorff, T.: The ethics of AI ethics: an evaluation of guidelines. Mind. Mach. **30**(1), 99–120 (2020). https://doi.org/10.1007/s11023-020-09517-8
13. Malik, T.: White paper on the use of artificial intelligence in trade facilitation (2023)
14. Jobin, A., Ienca, M., Vayena, E.: The global landscape of AI ethics guidelines. Nat. Mach. Intell. **1**(9), 389–399 (2019). https://doi.org/10.1038/s42256-019-0088-2
15. Hagendorff, T.: Blind spots in AI ethics. AI Ethics (2021). https://doi.org/10.1007/s43681-021-00122-8
16. Gillespie, N., Curtis, C., Bianchi, R., Akbari, A., van Fentener Vlissingen, R.: Achieving trustworthy AI: a model for trustworthy artificial intelligence, Australia (2020). Accessed 11 Apr 2022
17. Gillespie, N., Lockey, S., Curtis, C.: Trust in artificial Intelligence: a five country study. Brisbane, Australia (2021). Accessed 10 May 2022
18. OECD: Recommendation of the council on artificial intelligence: artificial intelligence policy observatory. OECD Legal Instruments (2019). Accessed 9 Nov 2021
19. Varshney, K.R.: Trustworthy Machine Learning (2022)
20. Thiebes, S., Lins, S., Sunyaev, A.: Trustworthy artificial intelligence. Electron Markets **31**(2), 447–464 (2020). https://doi.org/10.1007/s12525-020-00441-4
21. HLEG on Artificial Intelligence: Ethics Guidelines for Trustworthy AI (2019)
22. Lechler, T.G., Thomas, J.L.: Examining new product development project termination decision quality at the portfolio level: Consequences of dysfunctional executive advocacy. Int. J. Project Manage. **33**(7), 1452–1463 (2015). https://doi.org/10.1016/j.ijproman.2015.04.001
23. Kratochwill, L., Richard, P., Mamel, S., Brey, M., Schätz, K.: Globale trends der künstlichen Intelligenz und deren Implikationen für die Energiewirtschaft: dena-ANALYSE (2020)
24. International Research Center for AI Ethics and Governance: The Ethical Norms for the New Generation Artificial Intelligence, China. Accessed 9 June 2022
25. HLEG on Artificial Intelligence: Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment (2020)
26. KPMG in the UK: Controlling AI (2019)
27. European Commission: Proposal for a Regulation of the European Parliament and the Council: Laying down harmonised rules on Artificial Intelligence (Artificial Inteligence Act) and amending certain Union Legislative acts (2021)
28. Blackman, R.: Ethical Machines: Your Concise Guide to Totally Unbiased, Transparent, and Respectful AI. Harvard Business Review Press, Boston (2022)
29. IEEE Standard Model Process for Addressing Ethical Concerns during System Design, IEEE SA. https://standards.ieee.org/ieee/7000/6781/
30. Martinsuo, M., Poskela, J.: Use of evaluation criteria and innovation performance in the front end of innovation*. J. Product Innov. Manag. **28**(6), 896–914 (2011). https://doi.org/10.1111/j.1540-5885.2011.00844.x
31. Bayer, F., Kühn, H.: Prozessmanagement für Experten: Impulse für aktuelle und wiederkehrende Themen. Springer Gabler, Heidelberg (2013). https://doi.org/10.1007/978-3-642-36995-7
32. Hayes, A.: BSI White Paper – Overview of standardization landscape in artificial intelligence (2019)
33. Bertelsmann Stiftung (HRSG.): From Principles to Practice - An interdisciplinary framework to operationalise AI ethics (2020)
34. ABB et al.: etami: aligning trustworthy and ethical AI between academia, industry, and society. https://www.etami.eu/etami.eu.html

35. Lukyanenko, R., Maass, W., Storey, V.C.: Trust in artificial intelligence: from a foundational trust framework to emerging research opportunities. Electron Markets **32**(4), 1993–2020 (2022). https://doi.org/10.1007/s12525-022-00605-4
36. Sharkov, G., Todorova, C., Varbanov, P.: Harnessing the power of responsible innovation: the shift towards human-centered skills and competences in AI engineering (2022)
37. Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., Srikumar, M.: Principled artificial intelligence: mapping consensus in ethical and rights-based approaches to principles for AI (2020)
38. Riel, A., Flatscher, M.: A design process approach to strategic production planning for industry 4.0. In: Stolfa, J., Stolfa, S., O'Connor, R. V., Messnarz, R. (eds.) EuroSPI 2017. CCIS, vol. 748, pp. 323–333. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-64218-5_27
39. Sjoerdsma, M., van Weele, A.J.: Managing supplier relationships in a new product development context. J. Purch. Supply Manag. **21**(3), 192–203 (2015). https://doi.org/10.1016/j.pursup.2015.05.002
40. Aloisi, A., de Stefano, V.: Between risk mitigation and labour rights enforcement: assessing the transatlantic race to govern AI-driven decision-making through a comparative lens. SSRN J. (2023). https://doi.org/10.2139/ssrn.4337517
41. Jacovi, A., Marasović, A., Miller, T., Goldberg, Y.: Formalizing trust in artificial intelligence. In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, Virtual Event Canada, pp. 624–635 (2021)
42. Lee, J., Moray, N.: Trust, control strategies and allocation of function in human-machine systems. Ergonomics **35**(10), 1243–1270 (1992)
43. Lazanyi, K., Maraczi, G.: Dispositional trust — do we trust autonomous cars?. In: 2017 IEEE 15th International Symposium on Intelligent Systems and Informatics (SISY), Subotica, Serbia, pp. 135–140 (2017)
44. Zerilli, J., Bhatt, U., Weller, A.: How transparency modulates trust in artificial intelligence. Patterns **3**(4), 100–455 (2022)
45. Dor, L.M.B., Coglianese, C.: Procurement as AI governance. IEEE Trans. Technol. Soc. **2**(4), 192–199 (2021). https://doi.org/10.1109/TTS.2021.3111764
46. Ponick, E., Wieczorek, G.: Artificial intelligence in governance, risk and compliance: results of a study on potentials for the application of artificial intelligence (AI) in governance, risk and compliance (GRC), December 2022
47. Raab, C.D.: Information privacy, impact assessment, and the place of ethics. Comput. Law Secur. Rev. **37**, 105–404 (2020). https://doi.org/10.1016/j.clsr.2020.105404
48. Lotlikar, P., Mohs, J.N.: Examining the role of artificial intelligence on modern auditing techniques. SMQ **9**(2) (2021). https://doi.org/10.15640/smq.v9n2a1
49. Koshiyama, A., et al.: Towards algorithm auditing: a survey on managing legal, ethical and technological risks of AI, ML and associated algorithms. SSRN J. (2021). https://doi.org/10.2139/ssrn.3778998
50. Flatscher, M., Riel, A., Kösler, T.: The need for a structured approach towards production technology roadmaps in innovation-driven industries. In: Barafort, B., O'Connor, R. V., Poth, A., Messnarz, R. (eds.) EuroSPI 2014. CCIS, vol. 425, pp. 251–261. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-662-43896-1_22
51. Millar, C.C.J.M., Groth, O., Mahon, J.F.: Management innovation in a VUCA world: challenges and recommendations. California Manag. Rev. **61**(1), 5–14 (2018). https://doi.org/10.1177/0008125618805111
52. Bennett, N., Lemoine, G.J.: What a difference a word makes: understanding threats to performance in a VUCA world. Bus. Horiz. **57**(3), 311–317 (2014). https://doi.org/10.1016/j.bushor.2014.01.001
53. Reinhardt, K.: Trust and trustworthiness in AI ethics. AI Ethics (2022). https://doi.org/10.1007/s43681-022-00200-5

54. Brattberg, E., Csernatoni, R., Rugova, V.: Europe and AI: leading, lagging behind, or carving its own way? (2020). Accessed 27 Apr 2023
55. Nowacka, A., Rzemieniak, M.: The impact of the VUCA environment on the digital competences of managers in the power industry. Energies **15**(1), 185 (2022). https://doi.org/10.3390/en15010185
56. Sari, R.P.: Integration of key performance indicator into the corporate strategic planning: case study at PT. Inti Luhur Fuja Abadi, Pasuruan, East Java, Indonesia. Agric. Agric. Sci. Procedia **3**, 121–126 (2015). https://doi.org/10.1016/j.aaspro.2015.01.024
57. Condliffe, J.: A Single autonomous car has a huge impact on alleviating traffic. MIT Technol. Rev. (2017). Accessed 27 Apr 2023
58. Emaminejad, N., Akhavian, R.: Trustworthy AI and robotics: implications for the AEC industry. Autom. Constr. **139**, 104298 (2022). https://doi.org/10.1016/j.autcon.2022.104298
59. Castro, D.: Europe will be left behind if it focuses on ethics and not keeping pace in AI development | view. Euronews (2019). Accessed 27 Apr 2023
60. Pries-Heje, J., Johansen, J., Messnarz, R.: SPI Manifesto (2010). https://conference.eurospi.net/images/proceedings/EuroSPI2012-ISBN-978-87-7398-154-2.pdf