AI and Ethics (2023) 3:355–367
https://doi.org/10.1007/s43681-022-00228-7

OPINION PAPER

The AI ethics maturity model: a holistic approach to advancing ethical data science in organizations

J. Krijger[1] · T. Thuis[2] · M. de Ruiter[1,2] · E. Ligthart[1,2] · I. Broekman[1,2]

Received: 12 August 2022 / Accepted: 7 October 2022 / Published online: 24 October 2022

Abstract
The field of AI ethics has advanced considerably over the past years, providing guidelines, principles, and technical solutions for enhancing the ethical development, deployment and usage of AI. However, there is still a clear need for research that facilitates the move from the 'what' of AI ethics to the 'how' of governance and operationalization. Although promising literature on the challenge of implementation is increasingly more common, so far no systemic analysis has been published that brings the various themes of operationalization together in a way that helps the gradual advancement of AI ethics procedures within organizations. In this opinion paper we therefore set out to provide a holistic maturity framework in the form of an AI ethics maturity model comprising six crucial dimensions for the operationalization of AI ethics within an organization. We contend that advancing AI ethics in practice is a multi-dimensional effort, as successful operationalization of ethics requires combined action on various dimensions. The model as presented is a preliminary result of literature analysis complemented with insights from several practical mutual learning sessions with some of the major public, private and research organizations of the Netherlands. The article contributes to the AI ethics literature and practice by synthesizing relevant aspects of operationalization and relating these to the praxis of AI in a maturity model that provides direction for organizations seeking to implement these ethical principles.

Keywords AI ethics · Maturity model · Operationalization · Organizational dimensions

1 Introduction

With the increased usage of artificial intelligence (AI), concerns regarding the ethical aspects of AI decision-making are growing. In response, the field of AI ethics has advanced considerably over the past years, providing guidelines, principles, and technical solutions for enhancing the ethical development and deployment of AI. However, these initial initiatives have met with increasing criticism. Critics point to the large gap between principles and practice, remarking that there still is a clear need for research that facilitates the move from the 'what' of AI ethics to the 'how' of governance and operationalization in organizations. Having good ethical principles, it is argued, is not enough to ensure ethical AI development and deployment.

This sentiment is not just shared among academics in the field of AI ethics. AI practitioners, both in private and public organizations, find that despite the availability of frameworks, a plethora of practical challenges remain before any of these standards can be implemented in their data science processes and procedures. For organizations seeking to implement AI ethics or willing to devote more effort to the operationalization of ethics in their data science practices there is relatively little actionable research that relates AI ethical principles to organizational praxis of AI. Although in the literature various relevant themes for the implementation and operationalization of AI ethics have been outlined, so far no systematic approach for the gradual advancement of AI ethics procedures within organizations has been published. In the current paper, therefore, we will provide an organizational AI ethics maturity model, where we outline the crucial elements in the operationalization of AI ethics from an organizational perspective as well as underline

✉ J. Krijger
krijger@esphil.eur.nl

✉ T. Thuis
thuis@rsm.nl

[1] Erasmus School of Philosophy, Erasmus University, Rotterdam, The Netherlands

[2] Rotterdam School of Management, Erasmus University, Rotterdam, The Netherlands

the steps that could be taken to advance data science ethics within an organization. Maturity models typically help organizations appraise how well they are doing and provide an evolution path towards a desired end stage. In the field of technology these maturity models describe to which extent organizations have mastered certain capabilities for optimal use of the technology. The current AI ethics maturity model similarly describes the extent to which organizations have mastered specific dimensions of the operationalization of AI ethics. The model as presented is a preliminary result of literature analysis complemented with insights from several practical Mutual Learning Sessions; interactive learning sessions with some of the major public, private and research organizations of the Netherlands. These sessions in particular provided some important insights in how AI ethics is operationalized and resulted in two strong guiding considerations for the draft of the maturity model as it is presented here.

First, it would improve the fit between theory and practice, and as a consequence might improve the uptake of the AI ethics maturity model, if the model takes existing governance structures and processes of both public and private organizations into account. For organizations to effectively attune to the potential ethical challenges and possible unintended negative consequences arising from the use of AI, it is not only efficient but also recommendable to develop their AI ethics policies and governance in close relation to existing review and AI development and processes already in place. For researchers working in the field AI ethics implementation, this entails a shift of attention from policies and principles towards meaningful controls, governance and the integration of tooling. As became apparent, while the research on theoretical accounts of the accountability and governance of AI is ample, research on how these insights can be related to existing governance practices or how to leverage existing committee structures are scarce.

Second, we found that, although the different guidelines and principles might give the impression that each principle (e.g. fairness, accountability or explainability) requires unique actions and can be realized separately or sequentially in relation to the other principles, this is seldom how operationalization works in practice. The operationalization of ethical principles for AI is not a process where each principle is developed in isolation of the others or as a unique trajectory. Looking at how organizations operate and how AI ethics relates to more general societal challenges organizations seek address, we would argue that AI ethics is a matter of 'ethical management' making the operationalization of AI ethics something multi-dimensional. Operationalization requires a rigorous and solid infrastructure for ethical evaluation that demands action on multiple dimensions at the same time, such as governance, policy and training. Reducing the operationalization of AI ethics to the implementation of tooling for fairness or the use of explainability dashboards in the workflow, will have limited effect if there is no guidance on how to interpret the dashboards, no clear view on how one relates the outcomes of fairness metrics to design decisions and no governance in place to evaluate these decisions. The AI ethics maturity model then, is structured around the idea that operationalizing these principles requires action from organizations on all the aspects of the AI ethics maturity model as, in isolation, neither of these aspects will have a significant impact on the actual implementation of ethics into data science processes.

In fact, for most organizations the uptake of AI ethics might be accelerated when, rather than coming at the topic of AI ethics as something novel, unrelated to existing processes and requiring new committees, roles or positions, they can align data ethics practices with their existing frameworks and procedures. That is not to say that all relevant aspects of data science ethics can or could be adequately addressed with existing expertise and within existing structures. On the contrary, AI ethics requires a very specific expertise, and the apt implementation of AI ethics will require fundamental changes in the development and governance processes most organizations have in place. The ability to fully benefit from this expertise, however, depends on an organizations ability to integrate it in, and align it with, existing processes.

These considerations result in an AI ethics maturity model that is geared towards promoting AI ethics adoption and towards the more efficient use of resources to achieve ethical data science practices. As progress on the maturity aspects tends to reinforce progress on the other aspects, we contend that motivating organizations to get started is more important than requiring them to develop a perfect action plan that describes how to bridge the gap towards an ideal desired state in great detail.

The article will proceed as follows: in the first section, the AI ethics literature will be reviewed for clues on the elemental aspects of an AI ethics maturity model and the notion of a maturity model will be discussed. As a maturity model must encapsulate all relevant aspects to the implementation of AI ethics within an organization, we set out to integrate many of the suggestions made in the literature, combined with insights from our Mutual Learning Sessions, into a single holistic model on an organizational level. Section two, then, will discuss our AI ethics maturity model. It will briefly introduce the elements we consider crucial to the operationalization of AI ethics on an organizational level and provide a brief sketch as of what the initial vs. the ideal state in organizations might look like. In the fourth section we will discuss the utility and applicability of the model for various contexts. In particular, we will look at how the model could fare in both public and private organizations and why it should be able to provide guidance for both. We will also discuss the questions and challenges remaining and

give an indication of the validation steps we intend to take in the upcoming year to empirically validate the presented model.

The article makes the following contributions to the literature: (1) it brings together and synthesizes the relevant aspects for the operationalization of AI ethics into one overview; (2) it relates these aspects to the praxis of data science development and deployment as it takes place within organizations; and (3) it provides a direction for organizations seeking to implement these ethical principles and gives them tools for maturity comparison that allows for efficient learning from each other.

2 Literature review and the aspects of AI ethics maturity

In reaction to the growing concerns with regards to the ethical and societal impacts of AI, guidelines for ethical AI development have been published by a large and diverse set of organizations [13]. The limitations of this principled approach have extensively been reviewed in detail elsewhere (e.g. [15]) and will not be discussed here. In general, the principled approach in AI ethics has been criticized for the ambiguity of the principles (Dignum and Theodorou [26]; [13], the lack of accountability mechanisms to enforce normative claims [11] and the missing infrastructure for practical implementation [18] have grown significantly in recent years. From an operationalization perspective, the ambiguousness of the proposed principles such as fairness and explainability hinders their implementation and might even hamper the development of effective accountability mechanisms to enforce these ethical codes (e.g. [3, 26]). Furthermore, there is a contextual and political dimension to how guidelines are interpreted and translated into practice. This subsequently results in a divergence in relevant ways the guidelines are to be implemented into organizational practices. As such, the principles might provide an answer to 'what' is needed, but give limited guidance to data scientists, decision makers or organizations in general on 'how' ethical AI should be operationalized. The gap between principles and practice, as it is often referred to [25, 19], results in limited uptake of the principles in the procedures of companies [28]. In their survey of industry practices Vakkuri et al. [27] found that developers perceive ethics as relevant but distant from the issues they face in their work. Given the lack of notable effects on industry practices different solutions are proposed to close this gap ranging from regulation and conformity assessments (e.g. EU AI Act) to auditing (Raji et al. [25]), and methodologies or strategies for ethical risk assessment (Floridi and Strait [10]).

However, these solutions either focus on only one part of the puzzle that is the organizational operationalization of ethical data science, or they provide limited guidance to organizations seeking to implement AI ethics in their processes and structures. To get a full overview of this puzzle as well as a clear understanding of how best practices can be developed, we argue that research on the practical implementation of AI ethics, as well as organizations seeking to operationalize AI ethics, could benefit from an AI ethics maturity model. As suggested by Vakkuri [29] a maturity model for AI ethics could help to, in addition to the frameworks, models, and other tools that are actively used in the field, make AI ethics principles more tangible.

Maturity models describe 'an anticipated, desired, or typical evolution path of these objects shaped as discrete stages" ([2] p. 213). Their underlying assumption is that organizational evolution will follow a predictable linear stage-by-stage path to a desired state. Felch et al. [9] distinguishes three purposes for maturity models: "they are adequate tools for (1) documenting the status quo, (2) developing a corporate vision for process excellence and providing guidance on that development path, and (3) comparing capabilities between business units and organizations" (Felch et al. [9], p. 5166). In other words, with maturity models companies can appraise their process maturity and find guidance for how to utilize resources and capabilities in alignment with maturity goals.

Used in technological contexts, the word maturity is often referred to in terms of "the state of being complete, perfect or ready" [14]. When it comes to the technical adoption of AI techniques, various of such maturity models have already been published. For example, Pringle and Zoller [21] propose a four core phased model where they distinguish AI Novice, AI Ready, AI Proficient and AI Advanced. Each phase indicates an improvement on the number of proactive steps taken, the integration with the strategy and the advanced understanding of AI applications. [7], combine insights from their research on Logistics 4.0 maturity models with AI maturity models to assess maturity levels in the areas of technology, innovation and product roadmaps. Another popular model often used in the industry is the Gartner AI maturity model, with levels ranging from 'aware' to 'transformational', where AI is increasingly becoming part of the 'DNA' of the organization.

As of yet, only a few promising attempts have been made to provide an advanced AI *ethics* maturity model. The most relevant and elaborate examples in this regard are a model released by Salesforce and a recently released model by Open Data Institute. To start with the former, Salesforce released a basic and simple maturity model for the ethics of data science in an organization. They provide four phases of maturity, (1) ad-hoc—(2) organized & repeatable—(3) managed & sustainable—(4) optimized & innovative, with each having their own characteristics and areas of attention. Although the model touches on aspects such as education,

bias assessment and mitigation tooling, policies and ethics reviews, attention for these aspects is restrained to a single step making it hard to track progress or to start certain developments within the organization. As such it provides levels of maturity but doesn't specify themes that are specifically developed throughout these phases. The ODI, in its data ethics maturity model [30] does a better job in this regard by proposing six themes combined with five maturity levels per theme. It thereby presents a model to assess and benchmark practices and cultures towards ethical data science in organizations along six dimensions:

(1) Organisational governance and internal oversight, which concerns the strategy and leadership responsibility around ethical data practices,
(2) Skills and knowledge, which highlights the steps required to create a culture where ethical data practices are embedded by identifying the knowledge sharing, training and learning required within an organization,
(3) Data management risk processes, which seeks to identify key business processes that underpin ethical collection, use and sharing of data, to identify and assess risks of harm,
(4) Funding and procurement, which focuses on the investing in ethical data practices and developing requirements for procurement,
(5) Stakeholder and staff engagement, which highlights the engagement with internal and external stakeholders and
(6) Legal standing and compliance, which addresses compliance with relevant laws, regulations and social norms.

On each of these dimensions organizations can go from 'initial' to 'optimising' with the specific levels being: initial (baseline), repeatable (refined and repeatable in individual teams and projects), defined (processes are standardized though not widely adopted), managed (widely adopted and monitored) and optimising (optimise and refine processes). The ODI model in this sense really incorporates the multi-dimensional aspect of ethical data science maturity. Without going into too much detail on the specific sublevels it provides a complete and technically accurate model for public organizations in particular. However, some limitations remain. For one, despite the intended use for benchmarking performance and supporting the development of an action plan, the model is mainly focused on the collection and use of data from an NGO/oversight agency perspective, not *data science* per se from an organizational perspective. In their article on data ethics and AI impact [15] discuss this relationship between data protection and AI impact in more detail and conclude that data ethics and AI ethics can't be used interchangeably. They denote different concerns and processes which will make "the move from data to AI ethics is unlikely to be straightforward […] when AI ethics built upon data ethics is applied practically" (p. 224). While organizational alignment between data ethics and AI ethics is of relevance for maturing data science ethics, we argue that current maturity models underemphasize the organizational praxis of improving structures and processes to ensure the responsible development and deployment of AI. Building on these insights from the literature the following section will outline how a draft of an AI ethics maturity model from an organizational perspective could be conceptualized and how the model presented here came about.

2.1 AI ethics maturity: selection of dimensions

The current AI ethics maturity model is the result of literature review as discussed above, but is also to a large extent rooted in insights gained from the consultation of public, private and research organizations in the Netherlands, with varying degrees of maturity in operationalizing ethics in their data science processes. More specifically, the model is the result of multiple Mutual Learning Sessions (MLS). These Mutual Learning Sessions are a particular form of Mutual Learning Exercises as they have been proposed by [31]. Mutual learning exercises (MLEs), as they have defined it, aim to bring together various groups of stakeholders (researchers, potential users, intermediaries, policy makers, professionals, students, media, broader publics) to facilitate an interactive learning process through mutual exposure of views and experiences, expectations and concerns. Discussing the impact of technologies through in-depth dialogues with expert participants allows for the mutual sharing of preliminary analyses and dilemmas. However, where MLEs seek innovative methods and forms of deliberation (e.g. by making the participant an active actor in an experimental deliberative performance) to go beyond the more traditional forms (e.g. panel discussions and lectures), our Mutual Learning Sessions comprised a simple structured small-scale session with experts participating on an invitation-only basis. Central to the sessions was the mutual sharing of specific data science ethics related challenges and developments. Experts with backgrounds varying from technical, managerial, legal to academic were invited to discuss preselected topics from their unique perspectives based on acquired experiences. The objective of the sessions is to evaluate use-cases or organizational bottlenecks presented by a participating organization to provide context to their current and future maturity level regarding AI ethics. In eight MLS's with major Dutch institutions we were able to collect insights from a broad range of participants with expertise ranging from financial services to telecommunications, aviation to ministries, and municipalities to social welfare institutions. As will be discussed in the Discussion

section, this mutual learning methodology also allows for further studying, advancing and validating the AI ethics maturity model in close partnerships with public, private and academic partners.

3 AI ethics maturity model

The discussed literature review and the MLS sessions resulted in a model with six dimensions as shown in Fig. 1: Awareness and Culture, Policy, Governance, Communication & Training, Development processes and Tooling.

For each dimension the model defines five levels indicating the maturity of the specific dimension. The levels can generally be characterized as follows:

- Level 1: first awareness about data science ethics is present among individual employees and related activities are being initiated.
- Level 2: orientation on frameworks, guidelines/principles, trainings on data science ethics takes place in a team or collective context.
- Level 3: context-specific ethical frameworks, guidelines/principles have been developed and are being implemented in data science processes.
- Level 4: safeguarding mechanisms for ethical data science have been set up and are integrated throughout the organization.
- Level 5: organization-wide integration, training, and monitoring on ethical aspects of data science applications in accordance with legislation and policy frameworks.

In the Sects. 3.1 to 3.5, the dimensions with corresponding levels of the AI ethics Maturity model are briefly discussed.

3.1 Awareness and culture

Awareness about AI ethical aspects, and a culture in which ethical practices are integrated in the organization is considered an important starting point for the operationalization of AI ethics. Where literature on training and raising awareness for AI ethics is plentiful, no systematic analysis has been done on how awareness for the ethics of AI evolves or can be evolved within an organization. The lack of research notwithstanding culture and awareness are important aspects of a growing maturity of an organization in AI ethics. Only when aspects such as governance and tooling are coupled with growing awareness of ethics in the organization can organizations advance in their AI ethics endeavours.

Awareness starts with individuals being aware of the debate around AI ethics and the ways in which it is impacting their day-to-day work [25]. As stressed by [5] individuals should be aware that AI is no substitution for a human ethical compass, and decisions made by algorithms are the result of human choices in earlier stages of the process. In a more mature organization, awareness is not solely focused on being familiar with the potential ethical risks but extends to ways in which ethical issues can be mitigated. This is where the awareness of ethics ties in with the introduction of ethics review boards, codes of ethics, and the engagement of stakeholders in data science processes [25]. Promoting a culture of ethics and creating awareness for the ethical aspects of data science development and deployment creates the crucial internal support to make progress on these other dimensions.

As is commonly seen, awareness often starts fragmented with individuals (Level 1) or a group of people (Level 2), sometimes from very different departments, taking an interest in the topic of AI ethics. As these people team up and join forces the awareness process is put in motion where more senior management is educated, and the first more formal efforts (workgroups, task forces, frameworks, and standards) can be initiated (Level 3). The further integration of initiatives throughout the organization could lead to organization-wide support and representative multidisciplinary groups working on ethical data science (Level 4). The highest level of awareness and a culture promoting ethical data science, requires the buy-in from senior, middle and junior management as well as broad and active involvement of different departments (business, technical, legal, compliance) in the organization (Level 5).

During the Mutual Learning Sessions, AI ethics experts from different types of organizations shared their experiences with the start and process of their ethical data science practices. Whereas organizations take slightly different routes, their paths converge to the path described above where, from a small topic that interests individuals, the theme grows to an organization-wide topic. This is also where more research is required as to determine, for example, how major hurdles such as management buy-in can be overcome. Also more research on the cultural dimension in relation to the success of ethical committees or councils and involvement of stakeholders within organizations could deserve merit. Organizations that were successful in their operationalization of AI ethics often cite the support from senior management that they have and are part of a larger organization-wide strategy on ethics (in data science).

3.2 Policy

With the plethora of guidelines and frameworks for ethical AI, Policy is often the most obvious starting point for many organizations. However, managers often feel the need to create specific company-internal ethical guidelines to

emphasize the principles. Using an existing code of ethics may not fully cover the specificity that is useful for ethical data science applications. Although getting started with agreeing on some organizational principles is the first step, it isn't always obvious how organizations can translate them into their own house of policies. Coeckelbergh [4] suggests that policy proposals concerning AI ethics often start from an ethical principle or fundamental right, such as the "no harm" principle or explicability. These principles are relevant for data science, as algorithms should avoid discrimination, manipulation, and should be auditable and understandable. Therefore, these key principles should form the basic consideration of any policy in relation to the development and use of data science to ensure ethical operations. Although policy documents are crucial for first awareness and focussing the discussion, an overemphasize on policy alone has limitations. [16] argue that ethical guidelines for AI development are currently often used as assurance for investors and the wider public, rather than as a successful tool for governance. This situation highlights the importance of a policy framework that not just outlines the relevant principles but also includes the specific steps or requirements to bring them into practice.

Organizations at the start of their maturity process will have minimal to no policy related to operationalizing ethics into data science (Level 1). In Level 2, there is a growing demand for policy on the ethical aspects of AI. Conversations have started and there is a first concept on overall principles. Based on several iterations, policy for ethical AI becomes available and is communicated to all relevant stakeholders. It is only after the appointment of a specific function or role in the organization tasked with the implementation and monitoring of the policy that Level 3 is reached. In a next phase, policy is widely implemented in most parts of the organization. Moreover, a central point is initiated where the policy is monitored, feedback is gathered and questions regarding the policy can be discussed (Level 4). Ultimately, a fully matured ethical data science process will have a dedicated ethical data science policy installed that explicates the most important fundamentals that are required when developing and using data science within the organization and translates them to specific processes and ways of working (Level 5).

In the MLS's, policy was one of the most discussed topics among organizations in the early stages of maturity. More specifically, organizations where exploring which department is responsible for making the policy draft, how to delineate responsibilities for the policy document and most importantly, how to ensure it doesn't remain 'toothless'. Organizations that have been successful in having an AI ethics policy in place often involved a very diverse group of internal stakeholders and consulted multiple related departments throughout the process of policy development. As [13] already found in the literature, also in practice the ethical principles included in policy documents converged towards five key themes: fairness, explainability, accountability, transparency and the 'human in the loop'. Explainability for these organizations was defined as the ability to explain the reasoning, functioning or outcomes of specific models towards different stakeholders whereas transparency was more the communication of data and model usage as such. All in all organizations perceived policy as a key stepping stone in furthering their AI ethics initiatives as it paves the way for the development of more elaborate governance structures to ensures the principles and goals are met.

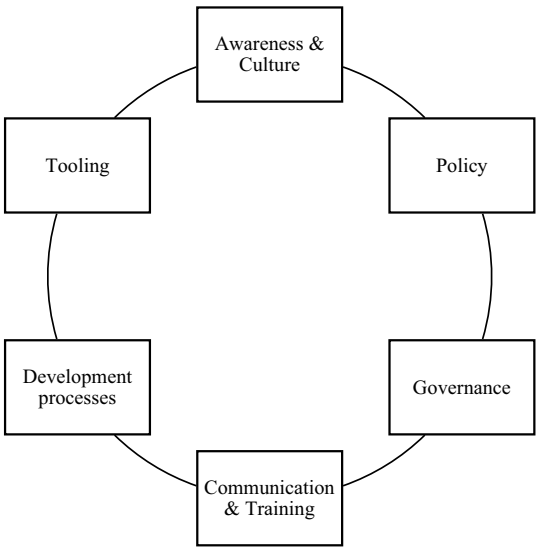

Fig. 1 AI ethics maturity dimensions

3.3 Governance

Governance is defined as the internal procedural ethical checks and balances in place in the development and deployment of AI systems. This might be in contrast with broader definitions used in guidelines such as the WHO's Ethics and Artificial Intelligence in Health (2021) where governance is seen as a political process that covers a range of steering and rule-making functions and involves balancing competing influences and demands. In the context of the AI ethics maturity model, governance is considered in accordance with corporate governance where governance is defined as the set of 'processes by which decisions relative to risk management and compliance are made within an organization' [17]. Firm governance is generally considered the structure of control within an organization. More specifically it concerns the control of agents within the organization, as governance processes are aimed at getting agents to act in accordance with the organization's interests [12].

Although governance is generally recognized as an important theme for the operationalization of AI ethics, so far few unified frameworks of AI governance have been published. Reddy et al. [24] point out that only few investigations have moved beyond the ethical aspects of 'what?' to consider the legal and governance aspects of 'how?'. This is partly due to a focus on technical solutions and assessment as well as limited attention and research on organizational contexts. To provide a five-level model of governance for the AI ethics maturity model, common practices in risk management and model risk management are utilized as indication for the minimally required steps organizations need to take towards adequate AI ethics.

An entry level form of governance entails the often already present data and/or privacy assessments and mandatory legal checks. However, broader model governance checks and evaluations have not been developed yet (Level 1). In a response to a demand for governance, ethical robustness and validation checks are introduced but not formally required in the organization (Level 2). Subsequently, these checks are being further developed and complemented with more specific ethical checks in design and post-hoc phases yet are still not formally required. Next to checks and procedures, initiatives for governance bodies concerning ethical AI are being introduced (Level 3). With the evolvement of AI applications, the level of standardization of governance procedures increases and ethical checks become formally required. Governance committees are appointed to oversee the ethical aspects of AI applications in the organization (Level 4). A mature organization consists of a fully integrated and supported AI ethics governance structure with formally required checks, procedures and operating governance committees (Level 5).

From our MLS's several practical aspects of initiating and developing ethical governance for data science and AI applications came to the fore. One trend that could be discerned was a growing mandate for the governance committees involved to ensure standards are adhered to and to enforce adoption of the ethical committee suggestions. In most organizations this mandate had to be acquired through various rounds of interdisciplinary consultation with various departments within the organization, moving from an advisory role at first to a formalized and authorized governance body in the more mature phases of development. This development ties in with a broadening scope for these governance bodies and further alignment: most organizations start out with two or three controversial cases that are widely discussed. Input from other committees or departments is collected and the space for new governance around the ethical aspects is explored. Once a decision-making procedure is put in place to address these specific concerns in these specific cases the scope starts to broaden: alignment with existing governance processes is improved and optimized. Participants from the financial sector stressed for example that the governance around AI ethics was carefully structured around the already existing committees such as the legally required Model Governance Committee, Privacy Office and the existing reviewing process for new services and products.

3.4 Communication and training

Internal communication about ethics as well as the training of data scientists and managers on the ethical aspects of AI is becoming increasingly crucial. Applying data science in decision-making processes involves communicating to stakeholders and the internal organization, where algorithms are used, how they are used and what the risks and implications are of these applications. Moreover, as Oliver and McNeil [20] note the responsible application of data science requires training in how to use data science and how to understand its impacts. However, they found little attention is paid to ethics in data science degree programs, suggesting that undergraduate data science degree programs may produce a workforce without the training and judgment necessary to apply data science methods responsibly. It is not surprising then, that the people with the skills to design and develop AI, deciding on issues in data collection, manipulation, and computation, have minimal training in performing their tasks ethically [25]. Next to designers, developers, and technical experts, also decision-makers need to be trained to consider the ethical implications of AI decision-making and solve the ethical dilemmas that emerge in the process [5]. Fortunately, there are increasingly more (online) courses, conferences, and training programs for all organizational departments available to improve knowledge levels throughout the organization.

Organizations at the start of their ethical data science maturity journey will have minimal communication and training options related to ethical data science. And if they do, communication and training are the result of an individual's interest and motivation (Level 1). In a further stage, ethical training would focus on a small group of key users, and communication primarily takes place in the core team working on data science applications (Level 2). Bringing together progress on the dimension of awareness and policy, the introduction of an ethical framework could help to establish a widespread understanding of ethics as well as the vocabulary needed for discussing issues related to data science ethics. The use of a framework enhances the communication and ethical training within the core team but also to key stakeholders outside the team (e.g. C-suite) (Level 3). Facilitating company-wide sessions on the topic of ethical data science, as well as the regular training of core team members in data science processes results in increasing the organizations communication and training maturity.

Communication about the ethical aspects is becoming a part of the daily tasks and activities of employees in data science (Level 4). A fully matured ethical data science organization will have fully adopted an ethical framework that helps establish a clear understanding of the vocabulary needed for discussing issues related to data science ethics. Communication about data science and ethical aspects is not only spread company-wide, but also happens outside of the company to customers and citizens. There will be a fully developed training module that includes a schedule for regular training for different types of users (Level 5).

For most participants in the Mutual Learning Sessions the final stages of communication and training maturity were still quite far away. Although some had progressed to regular internal communications about data science ethics and made AI ethics a recurring theme in modelling and AI team meetings, regular training of employees was not on the agenda anytime soon. Although there is a desire to start training programs and encourage data scientists to take part in ethics programs it became clear that buy in from senior management was an important barrier to actually getting both the communication and the training of the ground. As it stands this transition, from level 2 to level 3, warrants more research and would benefit from further insights in how to make training programs common practice in organizations.

3.5 Development processes

AI as we see in many organizations is going through different data science stages, design, development, testing, and deployment, before it is being used in day-to-day decision-making. Models such as the CRISP-DM cycle [22] are used to show the distinct tasks and focus points during each phase of the data science lifecycle. Ethical data science maturity requires the integration of ethics in the different data science lifecycle stages within the organization. The integration of ethics can stretch from the enforcement of standards in development practices, go/no go decision moments, standardized workflows, and the involvement of ethical boards and stakeholder groups. As ethical data science can have distinct implications in each of the phases, determining the required actions per stage is considered essential for addressing the (potential) ethical issues emerging over time.

In terms of development processes, a starting organization overall is lacking a structural approach to developing models. Consequently, ethics is not yet or only considered on an incidental basis (Level 1). Initiatives for a more structured approach to data science often lead to opportunities for making considered technical but also ethical choices (Level 2). The presence of a relatively structured data science approach in the organization then enables the implementation of ethics in specific parts of the lifecycle. Think for instance about the trade-off between interpretability and performance in the design phase, or the check for bias in the data in development (Level 3). While the consideration of ethical aspects in different phases is signalling a growing maturity, aligning ethical data science activities is crucial for addressing issues related to multiple phases in the cycle (Level 4). A mature level is reached in case of the integration of ethics in the entire data science workflow where specific activities are not only implemented in distinct lifecycle phases but also aligned throughout the cycle if required (Level 5).

Insights from our MLS indicate that structured approaches towards ethical data science are growing, and even perceived as essential by experts. Strategies for the integration ethics in the data science workflow are more and more developed, yet the actual implementation in development practices is considered challenging. One of the factors that is largely influencing the integration of ethics in development processes is the level of data science maturity of the organization. In the end, a structured data science approach allows for a more structured integration of ethical aspects in the different stage of the AI lifecycle. Additionally, the context and working conditions of AI developers could potentially influence the time required to go through various ethical checks and procedures. Overall, one of the major takeaways of the MLS regarding development processes is that successful adoption of AI ethics cannot be realized without integration in the development processes themselves. AI ethics is not just a suite of top-down policy and control measures but needs to be integrated in development processes, project management documentation and in the onboarding and handing-off of AI projects within organizations. Integration in these processes will make adoption of AI ethics measures and additional checks and balances easier and more time efficient.

3.6 Tooling

Eitel-Porter [6] addresses the importance of tooling to effectively implement AI ethics in a company's existing policies and governance structures. The dimension of Tooling here is defined as any (technical) method or tool that is used by organizations to implement ethical policies for AI. Think for example of dashboarding on fairness metrics, explainability functionalities, tools facilitating standardized workflows, tools for governing practices (monitoring performance AI, monitoring governance mechanisms (who did what at what point) or a platform to post/evaluate use cases on ethical aspects). As it has been remarked by Ayling and Chapman [1], this is a domain that has been left behind in the AI ethics gold rush of the past few years. In their recent review on proposed tools to operationalise ethical principles for AI they found four main groups of tools: (1) Impact Assessments, (2) Audits, (3) Participatory methods, and (4) Technical and design tools. Impact Assessments assess the impact of some

X upon some Y where, in AI ethics, X is often a specific AI solution and Y is often related to societal, environmental or privacy related impact. Similar to the more established Privacy Impact Assessment they provide a checklist to predict unintended negative consequences of technical innovations and to address stakeholder concerns. An audit consists of the examination of evidence of a process or activity, in the case of AI ethics the engineering process, and then evaluation of the evidence against some standards or metrics, which could be a regulation or AI ethics policy. Thirdly, although underdeveloped, participatory methods, such as the Delphi method, for the operationalization of ethics in AI are becoming increasingly more popular. Generally speaking these methods seek to involve stakeholders in the production and deployment of new technologies. The last category of tooling has received the most attention over the past years and comprises a range of computational approaches and quantitative metrics for ethical principles such as fairness and explainability. This form of tooling seeks to provide insights through dashboarding and provides methods to 'debias' training data sets pre- in- or post processing and to provide local or global explanations of 'black box' algorithms. fairness and explainability. This form of tooling seeks to provide insights through dashboarding and provides methods to 'debias' training data sets pre- in- or post processing and to provide local or global explanations of 'black box' algorithms.

In our maturity model we define the maturity starting point regarding tooling as no or minimal tooling being used in the context of operationalizing ethics into AI practice (Level 1). As the demand for gaining insights into the ethical implications of AI increases, ideas are gathered and translated into propositions for new methods and tooling (Level 2). With the advancement of these propositions, first methods and tools are implemented and adopted to gain insights into the ethical aspects of AI in the organization (Level 3). In a later stage, tooling is frequently used for monitoring, discussing, and improving AI ethics in different parts of the organization. At this point, the tooling is available for, and adopted by, multiple stakeholders (Level 4). Ultimately tooling should advance into a phase of proactive and continuous monitoring of ethical impact where both internal and external stakeholders are involved and use the available tooling to monitor, discuss, and improve ethical data science aspects (Level 5).

In our MLS it was stressed that, although dashboarding and other 'quick technological fixes' are for data scientists and innovation developers some of the easiest tools to adopt, their actual embedding in development processes prove quite difficult. Dashboarding fairness or explainability is valuable in setting off the discussion and, with open-source tools abounding, can be introduced in an early level. However, integrating the results in existing reporting and developing effective decision-making structures around the findings requires more advanced governance and policy structures to be in place. If these are not co-developed with the implementation of tooling data scientists remain with the question of 'what to do?' after they've found indications of bias or possible explainability issues. It could therefore be said that the successful integration of tooling is, similar to the other dimensions, very much dependent on progress of the other dimensions. Participants indicated that tools developed within the organization had a much higher adoption rate than external tools, checklists and programs that were often used only once. Indeed, it seems there is somewhat of a counter intuitive paradox in the domain of tooling we describe as the 'readiness-embeddedness paradox': as tools for ethical AI become more accessible and easier to use the ease with which they can and will be embedded in actual processes and practices seems to decline.

To assess the applicability of the maturity model a brief example case study is warranted. Suppose within an organization, Organization A, there is some awareness around the relevance of ethics for data science processes and the first working groups on the topic have been formed. The working groups opted for quick and ready to buy technological solutions for many of the ethical challenges resulting in the wide adoption of tooling such as explainability tools and fairness dashboards. Although the tools have been neatly integrated in the development processes the organization now faces a couple of challenges as both data scientists and senior management raise questions about the use of the tooling. As bias is found in some of their models they find themselves ill equipped to address these issues with no formal structure or framework to help guide the decision making process and no committee or group to vouch for the decisions. Mapping this organization on the maturity model would indicate that although in the dimensions of tooling and development processes, and to a lesser extent awareness, significant steps have been made already. However, based on the maturity model, it could be recommended that the organization should focus its resources on developing and implementing a policy framework, together with a team or department responsible for the policy where questions can be directed towards, as well as on getting a governance framework off the ground that can facilitate the decision making around the dilemmas that arise from the use of tooling. The above dimensions and maturity levels result in the following maturity diagram (Table 1).

4 Discussion

Now that a draft of the AI ethics Maturity models has been discussed in this final section we will discuss some considerations regarding the utility and applicability aspects of

Table 1 Ethical data science maturity overview

	Level 1	Level 2	Level 3	Level 4	Level 5
Awareness & culture	Awareness of data on an individual level out of personal interest	Fragmented attention throughout the organization	Focused and synthesized awareness through the formation of specific working groups or task forces	Organization wide support and representative multidisciplinary working groups	Buy-in from senior, middle and junior management, broad support and active involvement of developers, business and management
Policy	Minimal to no policy available for warranting ethics in data science	There is a demand for policy. Conversations have started and there is a first concept on the policy	Policy for ethical data science is available. A person assigned for the implementation and monitoring of the policy aspects	Policy is implemented in most parts of the organization. A central point is initiated for questions, monitoring, and feedback	Policy on data science ethics is widely implemented and monitored throughout the organization
Governance	Only legally mandatory checks	Additional robustness and model validation checks, not formally required	Specific ethical checks in design phase or post hoc, not formally required	Formally required ethical checks throughout data science lifecycle, governance committees are appointed	Fully integrated and supported AI ethics governance structure with formally required checks, procedures, and operating governance committees
Communication & Training	Minimal to no communication; employees improve their understanding based on own initiatives	Initiatives for training and communication only in small teams involved in data science processes	Incorporation of training and communication not only inside data science teams but also key stakeholders (e.g. C-suite) in line with established ethical framework	Company-wide sessions as well as the regular training of core team members Communication about the ethical aspects is becoming a part of the daily tasks and activities	Communication happens outside of the company to customers and citizens. There is a fully developed training module that includes a schedule for regular training for different types of users in the organization
Development processes	No structural approach to data science, or ethics in the lifecycle phases	Initiative for a structured data science approach mainly focusing on technical design choices in the development process	Relatively structured data science approach with ethical design choices were requested (on demand)	Structured approach, with alignment of ethical data science aspect to different phases in the data science lifecycle	Integration in the entire data science workflow where specific activities are implemented in and aligned with distinct lifecycle phases
Tooling	No or minimal tooling is used	There is demand for insights into the ethical aspects of data science. First ideas are gathered and translated into possible analysis/tooling	First methods and tools for generating insights into the ethical aspects are implemented and adopted	Tooling is available for and adopted by multiple stakeholders in the organization to monitor, discuss, and improve ethical data science aspects	Wide adoption of tooling where both internal and external stakeholders are using the available tooling to proactively monitor, discuss, and improve ethical data science aspects

the model for various contexts and outline future research directions as well as our intended steps for further validation of the model.

One initial concern with the development of the AI ethics maturity model was how it could fare in both public and private organizations and whether one general model could provide guidance for both. However, it became apparent that, although their organizational goals and values might differ, the challenges of operationalizing AI ethics were remarkably similar. All organizations were working on one or more dimensions of the model and confirmed the relevance of the other dimensions for their organization. As we seek to provide all organizations with a model help them operationalize ethical guidelines into their practices, we carefully selected and discussed the dimensions and levels as they should be applicable to different types of organizations. By keeping the steps intentionally very general there is a risk that the levels or dimensions might not perfectly fit with every industry or sector. However, we would argue that the benefit of providing general guidance on the themes and steps organizations could take outweighs that risk. In this way, the diagram is intended to guide the organization in the enhancement and optimization of ethical data science from not only a technical but also an organizational perspective. We acknowledge, of course, that despite the applicability of the diagram to various organizations, its implementation is highly dependent on the context of the organization at hand. Depending on its sector, industry, type of data, and decisions, the dimensions and levels in the diagram can have different implications.

When for instance considering the difference between a private and public organization, one of the areas potentially influencing the implementation of the maturity diagram are the transparency requirements. As a public organization, there is an obligation to be able to share publicly how decisions are made, why they are made, and what type of implications they have on citizens. This also applies to the context of algorithms, where restrictions are posed to what type of algorithms can be explained and thus utilized. Whilst private organizations might have a similar objective of being transparent to customers about the use of algorithms within their organization, the transparency is bounded by the potential loss in competitive advantage. The algorithmic model is considered as intellectual property by the organization and when shared publicly it can affect both market share and profit. For public organizations this trade-off is less problematic as public organizations frequently collaborate to further enhance the use of algorithms within the public domain. These differences can be traced back to the intended purpose for algorithms in general. In a private organization, profit-maximizing is driving its day-to-day operations, and with increased efficiency and effectiveness through algorithms this is to be further enhanced. In public organizations, the use of algorithms are aimed at benefitting the common good. Notice however that despite the differences in purpose or function, and the subsequent differences when implementing policies and governance mechanisms as suggested in the maturity diagram, the main categories still remain relevant for each organization. Although different trade-offs might arise depending on sector or industry, what is required on an organizational level to successfully operationalize ethics in AI seems remarkably similar. Rather, it seems, it is the risk level of an industry that determines the concretization of the maturity levels.

The application of tooling for example, such as a fairness dashboard, is less applicable to a spam filter than to a fraud detection or credit scoring model. Moreover, an organization deploying hundreds, or thousands of AI models will be more in need of standardized workflows, checks and balances, and AI ethics awareness throughout the organization than an organization that uses one or two models. Organizations also can also differ in the general knowledge levels about AI and its (ethical) consequences. Hence, a small organization consisting mainly of data scientists, ML engineers and analyst will require other communication and training activities than an organization with employees with high domain knowledge and low algorithmic experience. A question that could arise here is to what extent the maturity diagram accounts for the difference in requirements of a context. Here we could interpret the model more dynamically: rather than each organization having to move through each step linearly, one could envision different starting and end levels per industry, organization type or organization size. While the elaboration and implementation of the diagram are likely to differ, research into the extent to which the diagram reflects the main steps an organization needs to undertake to be ethically mature in the field of data science would be a valuable next step.

Other avenues for further research can be envisioned. One possible avenue relates the maturity levels to the field of standardization. The ideal (level 5) end state on one or all of the dimensions as they are described in the model could translated into industry standards or best practices. For example, the model could help to determine per industry what, for example, would be the minimally required level of maturity per dimension. This would help raise industry standards for AI and Ethics and provide a first starting point of sectors to translate maturity levels to relevant steps for their sector to become more mature. The most important avenue will be the general validation of the model. We set out to continue the Mutual Learning Sessions that helped in drafting the model, leaning heavily on a widely diverse group of experts, to assess the above assumptions as well as the effectiveness of the diagram in organizations across sectors. The upcoming year the diagram will be evaluated in a new series of sessions with experts and will be applied to different types of organizations. The MLS's will be used to

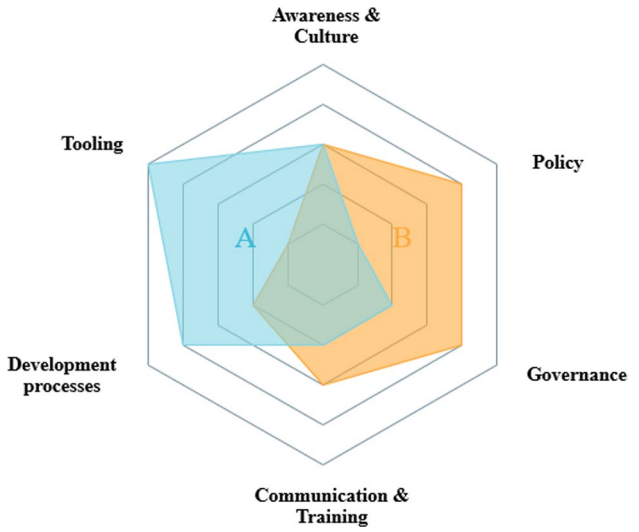

Fig. 2 Ethical Data Science Maturity Mapping–This figure shows the comparison between two fictional organizations and their maturity levels in the different dimensions. Organization A, as described in the earlier case study, has achieved high maturity levels in their tooling and development processes but is less mature in on other dimensions such as policy, governance, and communication. Organization B has established maturity policy and governance levels yet has lower maturity levels in the development processes and tooling dimensions

cross-validate the adequacy and applicability of dimensions as well as the generalizability of the maturity levels presented in the model. Furthermore, by mapping the maturity of actual organizations on the model, we seek to assess the action-guiding aspect of the model by formulating concrete action plans from the current state of AI ethics to a desired level of maturity in its context. By then following the maturity of the specific organizations over time, the diagram can be evaluated and further enhanced based on the process. We therefore made the model in way that allows for comparison between organizations but also compare an organization over time. Figure 2 displays how two organizations can be mapped with respect to the AI ethics Maturity dimensions and differ significantly in its focus. These forms of mapping provide insight in how knowledge between organizations can be shared effectively and where possibilities for cooperation can be found. As the goal of ethical AI development and deployment is widely and commonly shared these forms of comparison can be an impetus for further mutual learning both within and between organizations.

As said, the validation of the model is an important next step and we welcome all academics and practitioners in the field of ethics and AI, regardless of their background or discipline, to engage with the diagram in the upcoming year. The proposed model could not only be of value to practitioners to map their organization and define next steps towards AI ethics maturity, but also allow research to study the operationalization of AI ethics from a more organizational perspective. Future research could for example empirically investigate how the implementation of the maturity levels differs among organizations, what the effect of the model is on an organization's actual AI ethics maturity levels, and what the relation is between AI ethical principles (fairness, accountability, transparency) and the dimensions in the model. Moreover, we expect that there are different routes that organizations can take to reach AI ethics maturity in their organizations depending on factors such as the type of algorithm, organization size, and industry. New studies could explore these routes and corresponding strategies to achieve AI ethics maturity in various organizations. These future studies can be both qualitative in the form of case studies or ethnographies and quantitative by conducting experiments or surveys in organizations. Further developing the model based on the outcomes could be the next step toward applying the model as a standardized way of reaching high AI ethics maturity levels among different types of organizations.

5 Conclusion

Given the difficulty for organizations to translate the plethora of ethical principles to practice, a holistic framework from an organizational perspective was one of the blatant omissions in the AI ethics literature so far. With the current draft for an AI ethics maturity model, we not only sought to address this shortcoming but also sought to provide a practical synthesis of relevant literature on the operationalization of AI ethics for organizations. In addition, we related the aspects of the maturity model to the praxis of data science development and deployment as it takes place within organizations, as to help organizations develop action plans or strategies for the implementation of AI ethics. Although the model still needs to be broadly validated and will undoubtedly have shortcomings it is our hope that the current draft can inspire researchers to advance the holistic organizational approach in AI ethics. Ultimately the responsible and beneficial use of Artificial Intelligence will depend on our capacity to bring AI ethics to the organizations that are developing and deploying these systems.

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article and have not received funding to assist with the preparation of this article.

References

1. Ayling, J., Chapman, A.: Putting AI ethics to work: are the tools fit for purpose? AI. Ethics. (2021). https://doi.org/10.1007/s43681-021-00084-x
2. Becker, J., Knackstedt, R., Pöppelbuß, J.: Developing maturity models for IT management. Bus. Inf. Syst. Eng. **1**(3), 213–222 (2009). https://doi.org/10.1007/s12599-009-0044-5
3. Crawford K, Dobbe R, DryerT, Fried G, Green B, Kaziunas E, Kak A, Mathur V, McElroy E, Sánchez AN, Raji D, Rankin JL, Richardson R, Schultz J, West SM, Whittaker M. AI Now 2019 Report. New York: AI Now Institute, 2019, https://ainowinstitute.org/AI_Now_2019_Report.html.
4. Coeckelbergh, M.: Artificial intelligence: some ethical issues and regulatory challenges. Technol. Regul. **2019**, 31–34 (2019)
5. De Cremer, D., Kasparov, G.: The ethical AI—paradox: why better technology needs more and not less human responsibility. AI. Ethics. **2**(1), 1–4 (2022). https://doi.org/10.1007/s43681-021-00075-y
6. Eitel-Porter, R.: Beyond the promise: implementing ethical AI. AI. Ethics. **1**(1), 73–80 (2021). https://doi.org/10.1007/s43681-020-00011-6
7. Ellefsen, A.P., Oleśków-Szłapka, J., Pawłowski, G., Toboła, A.: Striving for excellence in AI implementation: AI maturity model framework and preliminary research results. LogForum. (2019). https://doi.org/10.17270/J.LOG.2019.354
8. Felch, et al.: Digitization in outbound logistics—application o.pdf. (n.d.). https://fis.uni-bamberg.de/bitstream/uniba/45549/1/VelchDigitizationse_A3b.pdf (2022). Accessed 5 Aug 2022
9. Felch, V., Asdecker, B., Sucky, E.: Digitization in outbound logistics—application of an industry 4.0 maturity model for the delivery process. In: Stentoft, J. (Ed.) Proceedings of the 30th Annual NOFOMA Conference: Relevant Logistics and Supply Chain Management Research. Kolding: Syddansk Universitet, pp. 113–128 (2018)
10. Floridi, L., Strait, A.: Ethical foresight analysis: what it is and why it is needed? Mind. Mach. **30**(1), 77–97 (2020). https://doi.org/10.1007/s11023-020-09521-y
11. 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
12. Haugh, T.: Harmonizing governance, risk management, and compliance through the paradigm of behavioral ethics risk. Uni. Pennsylvania. J. Bus. Law. **21**(4), 873 (2019)
13. Jobin, A., Ienca, M., Vayena, E.: The global landscape of AI ethics guidelines. Nature. Machine. Intelligence. **1**(9), 389–399 (2019). https://doi.org/10.1038/s42256-019-0088-2
14. Kärkkäinen, H., Myllärniemi, J., Okkonen, J., Silventoinen, A.: Maturity assessment for implementing and using product lifecycle management in project-oriented engineering companies. Int. J. Elect. Bus. **11**, 176–198 (2014). https://doi.org/10.1504/IJEB.2014.060218
15. Kazim, E., Koshiyama, A.S.: A high-level overview of AI ethics. Patterns. (2021). https://doi.org/10.1016/j.patter.2021.100314
16. Kerr, A., Barry, M., Kelleher, J.D.: Expectations of artificial intelligence and the performativity of ethics: implications for communication governance. Big. Data. Soc. **7**(1), 2053951720915939 (2020). https://doi.org/10.1177/2053951720915939
17. Miller, G.P.: The Law of Governance, Risk Management, and Compliance. Wolters Kluwer, Alphen aan den Rijn, Netherlands (2014)
18. Mittelstadt, B.: Principles alone cannot guarantee ethical AI. Nat. Mach. Intell. **1**(11), 501–507 (2019). https://doi.org/10.1038/s42256-019-0114-4
19. Morley, J., Elhalal, A., Garcia, F., Kinsey, L., Mokander, J., Floridi, L.: Ethics as a service: a pragmatic operationalisation of AI ethics. Minds. Machines. **31**(2), 239–256 (2021)
20. Oliver, J.C., McNeil, T.: Undergraduate data science degrees emphasize computer science and statistics but fall short in ethics training and domain-specific context. PeerJ. Comp. Sci. (2021). https://doi.org/10.7717/peerj-cs.441
21. Pringle, T., & Zoller, E. An AI maturity assessment model and road map for CSPs. 18. 2018
22. Provost, F., Fawcett, T.: Data Science for Business: What you need to know about data mining and data-analytic thinking. O'Reilly Media, Inc., Sebastopol, California, United States (2013)
23. Raji, I.D., Smart, A., White, R.N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., Barnes, P.: Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing. In: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20). Association for Computing Machinery, New York, pp. 33–44 (2020). https://doi.org/10.1145/3351095.3372873
24. Reddy, S., Allan, S., Coghlan, S., Cooper, P.: A governance model for the application of AI in health care. J. Am. Med. Inf. Assoc. **27** (2019). https://doi.org/10.1093/jamia/ocz192
25. Stahl, B.C., Antoniou, J., Ryan, M., Macnish, K., Jiya, T.: Organisational responses to the ethical issues of artificial intelligence. AI. Soc. (2021). https://doi.org/10.1007/s00146-021-01148-6
26. Theodorou, A., Dignum, V.: Towards ethical and socio-legal governance in AI. Nat. Mach. Intell. **2**, 10–12 (2020). https://doi.org/10.1038/s42256-019-0136-y
27. Vakkuri, V., Kemell, K.-K., Kultanen, J., Abrahamsson, P.: The current state of industrial practice in artificial intelligence ethics. IEEE Softw. **37**(4), 50–57 (2020). https://doi.org/10.1109/MS.2020.2985621
28. Vakkuri, V., Kemell, K.-K., Kultanen, J., Siponen, M., and Abrahamsson, P. (2019). Ethically aligned design of autonomous systems: industry viewpoint and an empirical study. *ArXiv*.
29. Vakkuri, V., Jantunen, M., Halme, E., Kemell, K.-K., Nguyen-Duc, A., Mikkonen, T., & Abrahamsson, P.: Time for AI (Ethics) maturity model is now. In: Espinoza, H., McDermid, J., Huang, X., Castillo-Effen, M., Chen, X.C., Hernandez-Orallo, J., OhEigeartaigh, S., Mallah, R. (eds.) SafeAI 2021: Proceedings of the 2021 Workshop on Artificial Intelligence Safety. RWTH Aachen. CEUR Workshop Proceedings, 2808 (2021). http://ceur-ws.org/Vol-2808/Paper_16.pdf
30. Yates, D., Maddison, J., Burton, J.: Data Ethics Maturity Model. https://theodi.org/article/data-ethics-maturity-model-benchmarking-your-approach-to-data-ethics/#:~:text=The%20data%20ethics%20maturity%20model%20is%20a%20tool%20for%20anyone,practices%20are%20across%20your%20organisation. (2022). Accessed 26 2022
31. Zwart, H., Brenninkmeijer, J., Eduard, P., Krabbenborg, L., Laursen, S., Revuelta, G., Toonders, W.: Reflection as a deliberative and distributed practice: assessing neuro-enhancement technologies via mutual learning exercises (MLEs). NanoEthics **11**(2), 127–138 (2017). https://doi.org/10.1007/s11569-017-0287-4