



Operationalising digital ethics: establishment of an ethics evaluation tool for data analytics

Jean Enno Charton¹ · Simon Lucas¹ · Werner Vogd² · Wolfgang Halter³ · Sarah Josefine Becker²

Received: 2 May 2025 / Accepted: 28 July 2025 / Published online: 29 September 2025
© The Author(s) 2025

Abstract

Digital ethics is increasingly applied in practice within contexts such as public service and corporate environments. However, a significant challenge remains in understanding how to implement these principles effectively without diluting their efficacy and meaningful impact. While organisations employ a range of solutions, such as committees, training, audits, or tools to bridge the often cited “principles to practice gap,” there is a recognised need to tailor these solutions to address domain-relevant challenges, such as in data analytics. Despite the substantial role of data analytics in driving decision-making processes and fostering innovation across various sectors, there is a notable gap in research focused on how digital ethics is operationalised in data analytics. Our work presented here addresses this gap by discussing how operationalisation of digital ethics can be achieved within this domain. The challenges and insights from the Analytics Center of Excellence at Merck KGaA, a multinational science and technology company, are utilised as a case study for implementing digital ethics within this domain. We present the development and deployment of Merck’s Digital Ethics Check (MDEC), an instrument tailored to address the ethical challenges in data analytics. The present work underscores the advantages of a semi-automated integration of digital ethics into data analytics. In doing so, we highlight the relevance of the challenges encountered in the practical application of digital ethics within the domain of data analytics in a business context. Our work provides a practical impetus to guide organisations seeking to embed ethical considerations within their data analytics operations.

Keywords Digital ethics · Operationalisation of ethics · Ethical guidelines · Data analytics · Digital Ethics Check

1 Introduction

As the digital transformation continues to accelerate across all levels of society, many organisations, including businesses, are seeking suitable ways to respond to associated

ethical challenges, such as bias, transparency, privacy and responsibility (Mittelstadt 2019; Rudschies et al. 2021; Kraus et al. 2021; Priyono et al. 2020; Díaz de la Cruz et al. 2025). Despite the development of various ethics codes and guidelines intended to navigate these challenges and bridge what has become known as the “principles to practice gap” (Corrêa et al. 2024; Georgieva et al. 2022; Mökander et al. 2023; Morley et al. 2020; Mittelstadt 2019), it has become apparent that these measures are not universally applicable solutions. In fact, a growing body of scholarship emphasises the necessity to adapt existing ethical principles and guidelines to better suit the distinct needs of differing institutions, people, projects, and workflows (Mittelstadt and Floridi 2016; Rochel and Evéquo 2021; Floridi et al. 2018; Jobin et al. 2019; Adams 2024; Amugongo et al. 2023; Krijger et al. 2023; Fjeld et al. 2020; Hagendorff 2020; Schiff et al. 2021; Floridi 2019; Morley et al. 2020; Hickok 2020; Giovanola and Tiribelli 2023).

Efforts to operationalise ethical guidance and close this “principles to practice gap” are ongoing at many levels. They

✉ Sarah Josefine Becker
sarah.becker@uni-wh.de

Jean Enno Charton
jean-enno.charton@merckgroup.com

Simon Lucas
simon.lucas@merckgroup.com

Werner Vogd
Werner.Vogd@uni-wh.de

Wolfgang Halter
wolfgang.halter@merckgroup.com

¹ Merck (Germany), Darmstadt, Germany

² Witten/Herdecke University, Witten, Germany

³ Merck Life Science KGaA, Darmstadt, Germany

have yielded a variety of approaches that range from specific checklists and exercises that teams and individuals can use to assess the digital ethics implications of their work (High-Level Expert Group on AI 2019; UN Global Pulse 2019; Ada Lovelace Institute 2022; UK Statistics Authority 2022; Deon 2018; Keller et al. 2020; Open Data Institute 2021; Epstein et al. 2018; d'Aquin et al. 2018; Saleiro et al. 2018; Chmielinski et al. 2020; Wachter et al. 2021; Zorio 2021; Meier et al. 2022), staff training modules (Danish Design Center 2021; Omidyar Network 2018) to review procedures involving dedicated ethics panels (Nemat et al. 2023; Schuett et al. 2023), and organisation-wide certification processes or ethics-based audits (Epstein et al. 2018; Saleiro et al. 2018; AI Ethics Impact Group 2020; Chmielinski et al. 2020; Cobbe et al. 2021; ForHumanity 2021; Poretschkin et al. 2021; Grellette 2022; Mökander et al. 2021, 2022; Zicari et al. 2021; Floridi et al. 2022; Mökander and Floridi 2021, 2022). Among these approaches, however, a common thread is emerging: the operationalisation of ethical guidelines is far from simple and is particularly challenging due to the complexity of the issues companies are dealing with.

Navigating the domain of data analytics, an essential component of any digital operation, exemplifies this complexity. Ethical concerns raised around data analytics encompass a range of issues. These include potential infringement of privacy, power imbalances and structural inequities, manipulation risks from exploiting people's vulnerabilities through data analytics, and biases that perpetuate disadvantages for specific groups (Martin 2022; Hirsch et al. 2021). In addition, quality issues and errors can lead to flawed results and predictions, and a lack of transparency and comprehensibility within data analytics processes can make it challenging to assess procedures and identify potential problems. These concerns increase especially around predictive analytics, where statistical patterns from historical data sets compiled from many people are used to make predictions about individuals, for example their consumer preferences or their likelihood to default on payments (Mühlhoff 2021).

While the literature has extensively explored the role of ethics codes, guidelines, and ethics panels as foundational elements in the operationalisation of digital ethics, the specific application of these ethical frameworks within the domain of data analytics remains understudied. Despite the critical importance of data analytics in driving decision-making and fostering innovation across various sectors, there is a significant lack of research on how established ethical frameworks are adapted and implemented specifically within analytics practices (Martin 2022; Mühlhoff 2021; Gotterbarn and Kreps 2021; Mittelstadt 2019; Hagendorff 2020).

In illustrating the application of existing research within our organisation as a means to bridge specific aspects of this gap, we herein describe how operationalisation can be effectively realised in practice specifically in data analytics

contexts (see also Zuber et al. 2022; Novak and Pavlicek 2021; Leonelli 2016; Ryan et al. 2024). We leverage the distinctive challenges and insights from the Analytics Center of Excellence at Merck KGaA (henceforth Merck), a multinational science and technology company, as an exemplary model for implementing digital ethics in this field. As a case study, we present the development and deployment of Merck's Digital Ethics Check (MDEC), a tool specifically designed to assist data scientists in addressing the ethical challenges inherent in data analytics endeavours within the sphere of Merck. We highlight the wide-ranging relevance of the challenges we encountered in practically applying digital ethics within the area of data analytics in a business context. Specifically, we identified three core challenges that hold broader implications for operationalising digital ethics within data analytics and discussed practical issues related to implementing tools to assess ethical risks in data analytics. We explore how this approach can be generalised and integrated into broader digital ethics implementation strategies, and how it contributes to the overall debate on operationalising digital ethics.

2 Operationalising digital ethics: challenges in data analytics

The multinational science and technology industry company Merck, specialises in healthcare, life sciences and electronics has dedicated teams and processes in place that are responsible for integrating ethical deliberation into its business practices for over a decade (Oschmann 2018). This is exemplified by the establishment of a dedicated Bioethics and Digital Ethics Office, entrusted with offering ethical guidance in contexts characterised by regulatory ambiguity and moral intricacy. Additionally, the Merck Ethics Advisory Panel for Science and Technology (MEAP), which supports the Bioethics and Digital Ethics Office with independent advice and consists of external academic experts from biology, medicine, bioethics, philosophy, and law, has been advising the company in ethically sensitive healthcare and life science areas such as stem cell research and genome editing (Sugarman et al. 2018).

To establish similar guidance for ethical questions arising throughout the company's ongoing digital transformation and digital business developments, Merck has created the Digital Ethics Advisory Panel (DEAP) with external experts from digital ethics, law, big data, digital health, data governance and patient advocacy (Merck 2021). The DEAP has developed a Code of Digital Ethics (CoDE), providing a structured framework that clearly delineates the relationships between principles and their specific applications to data or algorithmic systems. As detailed elsewhere (Becker et al. 2022), the principles are, on the one hand, formulated

at a level of abstraction that allows for their application across various domains relevant to data and AI ethics, while also being designed to be practically implementable. Each core and subsidiary principle within the CoDE is precisely mapped to its level of relevance, specifying whether it pertains to data or algorithmic usage. This mapping is further elaborated with detailed guidelines, enriched by tangible examples. These examples demonstrate potential violations or neglect of the principles in practical environments involving data or algorithm use and propose exemplary solutions. For instance, the subsidiary principle of “explainability,” which is part of the core principle of autonomy, pertains specifically to algorithms. It articulates a clear directive or commitment to the explanation of algorithmic systems. This principle exemplifies issues that may arise such as those related to automated decision-making and the subsequent impact on individuals. The CoDE guides the advisory work of the DEAP (Nemat et al. 2023), provides guidance to employees, serve as an external affirmation of quality to build trust, and lead the company’s wider efforts to operationalise digital ethics. These include, for example, training all employees in the basics of digital ethics as well as more advanced workshops for teams working with data and algorithms.

While Merck’s creation of the Bioethics and Digital Ethics Office and the formation of the MEAP and the DEAP constitutes important mechanisms for facilitating ethical deliberation, these efforts alone have proven not to be adequate for dealing with digital ethics specifically within data analytics. The reasons underlying this limitation arise from the specific and complex challenges inherent in data analytics—and potentially other applications of AI—when examined from both practical and ethical perspectives. For a better understanding of these challenges, in this section we will examine the practical operations of Merck’s Life Sciences Analytics Center of Excellence (ACE). Through this exploration, we identify three important challenges encountered when translating digital ethics into practice.

Essentially, ACE drives the Life Science unit’s global data and analytics strategy and acts as a service unit. The focus is on providing enterprise-wide oversight and delivery of secure, high quality, timely data and applications that support data-driven decision-making and thinking. The ACE portfolio currently includes over 1000 use cases that are in constant development and managed and processed centrally via a data integration platform.

The use cases managed by ACE are diverse and can address issues such as process optimisation, quality improvement or cost savings. An example is the development of the COMPASS tool, a data platform that supports sales teams to find the right customer at the right time. Based on the analysis of large amounts of complex data from previous customer interactions, the tools use machine learning to

predict which potential customers are looking for business partners to outsource some of their development and manufacturing processes.

As ACE develops agile and innovative digital services, the Digital Ethics Office has encountered an increasing and substantial volume of ethical inquiries arising from this division. These concerns include the sensitivity of collected customer information, the awareness of customers about being targeted by data-driven tools, and the biases that may result in certain customer segments being targeted unfairly. The frequent engagement with ethical questions that proved too complex to be resolved within the ACE team underscored the need for collaboration with Merck’s Digital Ethics Office. This realisation emphasised the necessity for a well-structured digital ethics framework, recognised by both management and staff at ACE. They specifically acknowledged that this framework should be seamlessly integrated into existing workflows to facilitate its effective implementation.

Initial discussions with the Digital Ethics Office have shown that the extensive automation within the project management framework at ACE necessitates an approach to ethical guidance that can be seamlessly integrated into this automated workflow. To effectively implement digital ethics in such a context, a methodological framework is required that offers swift and readily accessible means for users to acquire dependable and contextually tailored ethical guidance for each specific use case. In our endeavours to evaluate ethical dimensions within ACE projects and given the unit’s project management approach, we confronted various challenges stemming from the inherent structure and configuration of the ACE unit, compounded by the distinctive hurdles encountered in the domain of data analytics.

2.1 Challenge 1: too many use cases

The first challenge arises from the overwhelming number of continuously developing use cases. With more than 1000 projects ongoing in ACE at any given time, the task of performing individualised ethics assessments by Merck’s Digital Ethics Office or eliciting in-depth advisory consultations by the DEAP becomes extraordinarily complex. For instance, the Office is staffed by only 3–4 individuals who consult for a workforce exceeding 60,000 employees, addressing requests that span biomedical research, data utilisation, and AI applications. Consequently, in-depth consultations can realistically be provided only for high-stakes cases. The scale of the operation of ACE hence necessitates exploring alternative mechanisms for ethical assessment, which may require integrating elements of automation to ensure thorough yet efficient evaluations (Mökander et al. 2023). These issues are not unique to ACE; they are common across the data analytics field and likely extend to other domains, including general AI applications. Many

data analytics centres handle multiple parallel projects, each requiring the processing and interpretation of large data sets from diverse, widely distributed sources, with varying data management requirements based on their origin. A key driver of this development is the increasing democratisation of analytics within organisations. (Leonelli 2016; Kroll 2018; Wang et al. 2019; De Bie et al. 2022). The high volume, velocity and variety characteristic of big data endeavours can make it difficult to quickly spot ethical issues such as bias or problems with data quality and validity (Leonelli 2016; Keller et al. 2020; Kroll 2018).

This identifies a critical requirement for operationalising digital ethics within such a setting. Specifically, a reliable system for evaluating use cases is required, like a risk-based categorisation method (Mökander et al. 2023). Such categorisation would ensure that cases demanding closer scrutiny are identified promptly and managed in a manner that maximises resource efficiency (Aiken 2021). This approach is imperative to keep pace with the volume of projects and ensure ethical considerations are not left by the wayside in the face of progress.

2.2 Challenge 2: complex and dynamic project management

Practical issues in ethics assessment also arise from the dynamic nature of the project management, where work during all phases uses incremental and iterative processes. This also applies to the data integration platform ACE is working on, which uses highly automated project management with consistent processes and workflows and a specific lifecycle concept for each use case. Each lifecycle consists of three phases: ideation and planning, proof-of-value development, and industrialisation. In the ideation phase, a project idea from a business unit is converted into a use case. It is checked for its data requirements, potential value, and implementation ease. Stakeholders collaborate to plan its implementation and evaluation. In the next stage, the proof-of-value phase, a first version of the use case is created and evaluated. If it is successful, it continues to the industrialisation phase where it is fully developed, supported, and launched.

In this project management lifecycle, frequent shifts in project strategy and objectives can introduce complexities in ethical evaluations, potentially leading to problematic instances of “scope creep,” wherein AI systems intended for specific contexts are deployed in divergent settings (Danks and London 2017; Aizaz et al. 2021). Moreover, alterations in project parameters, such as changes in methodology or the nature of utilised data, such as transitioning from using non-personal to personal data, can further complicate ethical inquiries. The dynamic nature of project oversight necessitates structured methods beyond one-time advisory

consultations, facilitating ongoing ethical assessments to address evolving circumstances effectively.

The issue of how to address ethical questions in highly dynamic project environments that involve some form of automation has also been discussed previously in other data analytics contexts (Zuber et al. 2022; Umbrello and Gambelin 2021). Data analytics often uses agile development and project management techniques and may consist of several iterations of sourcing, preparing, and integrating data, creating variables and profiles, and building, training and refining models (Wirtz et al. 2023; Keller et al. 2020). Similarly, data analytics often entails the repurposing and recontextualisation of data, potentially influencing project impacts in context-specific manners and complicating the prediction of their long-term ethical implications. (Kroll 2018; Keller et al. 2020; Deutscher Ethikrat 2017). Such complex workflows and project life cycles therefore necessitate careful analysis and design for how digital ethics can be integrated into different steps of data analytics development and deployment process (Hirsch et al. 2021; Bleher and Braun 2023; Umbrello 2020; Umbrello and Gambelin 2021; Zuber et al. 2022).

2.3 Challenge 3: limited ethical expertise

Given the varied composition of team structures in ACE, an additional challenge emerges. While teams exhibit a high degree of heterogeneity, comprising both highly specialised experts and generalists, there is a notable scarcity of expertise in the domain of data and AI ethics. Ethical deliberations in practice often rely on subjective individual assessments and are influenced by the personal sensibilities of involved parties. Assessments and consensus within teams are further complicated by various factors, including conflicting interests such as a project owners’ inherent inclination towards project progression. Moreover, teams frequently operate across different geographical locations, where moral decisions must navigate conflicting local norms and sensitivities.

Additionally, ambiguity occasionally already arises regarding what even constitutes an ethical question, with many questions primarily revolving around legal considerations such as data protection or trade compliance. Furthermore, within these diverse teams, determining accountability for project outcomes, particularly in terms of ethics, is challenging. The dispersed nature of data analytics can obscure the detection of ethical quandaries and the assignment of accountabilities, particularly when individual choices or behaviours seem ethically insignificant. If such allocations are unclear, there is a risk that responsibility will be diffused, leading to a situation where no one takes ownership of the issue.

Such questions surrounding capacity building and accountability hold widespread significance within the realm of digital ethics. Numerous proposals have emerged aiming to empower non-ethicists in addressing moral inquiries pertaining to data and algorithms, while also elucidating roles and responsibilities amidst organisational intricacies. For example, given the complexity of data analytics and its interdependence with varied partners, it has been posited that an adequate infrastructure to support the reliable implementation of digital ethics across distributed systems is necessary. According to Luciano Floridi, such an ethical infrastructure, or “infraethics,” is especially important where small actions by many people may combine to create large morally good or bad outcomes (Floridi 2012). The distributed nature of data analytics can create difficulties in spotting ethical issues and assigning responsibilities, especially when every small individual decision or action may appear morally negligible (Leonelli 2016). Where such assignments are lacking, there is the risk that “everybody’s problem becomes nobody’s responsibility” (Floridi and Sanders 2004). By sensitising participating individuals to ethical issues and providing ethical guidance to facilitate morally favourable results even in highly distributed systems, a suitable ethics infrastructure can also improve clarity and consistency of morally relevant decisions. Without such guidance on the understanding, management, and mitigation of digital ethics issues (Kelley 2022; Wirtz et al. 2023) data analysts and managers alike may be more likely to base ethical decisions on “informal benchmarks” rather than relate them to clear principles (Hirsch et al. 2021), and the extent to which the experience, expectations and interactions of team members shape development processes may go unnoticed (Passi and Sengers 2020).

In summary, operationalising digital ethics in the field of data analytics reveals multiple challenges, each highlighting the need for a more nuanced and integrative approach to the subject. The first complication arises from the sheer volume of use cases, which poses significant difficulties for thorough and individualised ethical assessments. The inherently high-speed, large-scale nature of digital enterprises calls for the development of efficient and perhaps automated mechanisms for ethical evaluation. Secondly, the dynamic nature of project management, particularly in highly automated environments, further complicates the integration of digital ethics. Ethical guidelines need to be adaptable and continuously appraised for them to remain relevant amidst swiftly shifting project strategies and objectives. Finally, the issue of capacity building remains critical. Ethical training needs to extend beyond one-off sessions to build a sustained collective consciousness about digital ethics.

3 Semi-automated digital ethics operationalisation in ACE

Supported by the aforementioned challenges, our decision to operationalise digital ethics within ACE has been directed towards the integration of a semi-automated tool into existing workflows. The adoption of semi-automated tools for operationalising digital ethics in organisational contexts is a response to the complex challenges faced by large and multifaceted entities and is gaining uptake in the academic literature (Zuber et al. 2022; Law et al. 2020; Kazim et al. 2021; AI Ethics Impact Group 2020). Alternative approaches, such as hiring additional ethicists, are less attractive for various reasons. Not only is this approach economically unviable, but it also offers less flexibility in responding to fluctuations in the volume of inquiries. Ultimately, it contradicts the goal of upskilling employees in ethics expertise to enable responsible bottom-up innovation without cumbersome external reviews.

A common theme that emerges from these proposals is that digital ethics should be embedded into the existing structures and workflows of data analytics departments. Some have suggested that this might best be achieved by strategically placing ethicists to collaborate with data and software professionals and other staff at key points (McLennan et al. 2022), others focus more on the procedural embedding of ethics checkpoints along a project’s pipeline from conception to development and implementation (Char et al. 2020) and tying these in with, for example, already established agile software development processes (Zuber et al. 2022). Automation of at least some aspects of digital ethics is also being explored, for example by integrating tools that check for issues such as bias or fairness into coding or data analysis environments (Chmielinski et al. 2020; Saleiro et al. 2018; Wachter et al. 2021) or by automatically generating risk scores after manual assessment of ethical aspects (AI Ethics Impact Group 2020). Even the use of machine learning itself to assess ethical issues has been suggested (Epstein et al. 2018; Meier et al. 2022).

However, both wholly technologically oriented and entirely human-centred methods fall short of addressing the complex requirements of digital ethics in data analytics (Leonelli 2016; Kazim et al. 2021; Mühlhoff 2021). Exclusively technical tools, while efficient and capable of processing vast amounts of data rapidly, frequently lack the capacity to meaningfully interpret ethical principles and navigate complex social contexts (Brown et al. 2021; Selbst et al. 2019). Such tools may struggle to adapt to the evolving nature of ethical standards and the diverse perspectives that characterise ethical issues in real-world scenarios. In addition, these tools are difficult to integrate

into a heterogeneous development platform such as that of ACE. They would need to be implemented in multiple locations, with the risk of creating blind spots. On the other hand, exclusively relying on human judgement for ethical decision-making, while invaluable for its depth of understanding and contextual awareness, faces limitations in scalability and consistency. Differing viewpoints and interpretations can lead to variability in ethical assessments and decisions, particularly in large organisations where consistent application of ethical standards is arguably crucial. This consistency is important, particularly to ensure that the ethical obligations and commitments enshrined in the company's values or principle frameworks are upheld by every employee, holding them accountable for acting in accordance with these principles. It is equally important to avoid the exploitation and manipulation of ethics as a corporate means to justify actions retrospectively, commonly referred to as "ethics shopping" (Floridi 2019). Additionally, this adherence to self-imposed commitments is prudentially beneficial, as it exemplifies a commitment to ethical responsibility and fosters trust among all of the company's stakeholders.

Therefore, we posit that a semi-automated integration of digital ethics into a data analytics department like ACE has advantages given the challenges outlined above. Many use cases in data analytics, particularly within the context of ACE, are ethically uncontentious. These use cases include projects such as developing simple graphical dashboards for data visualisation. Automatically filtering out such straightforward projects enables an organisation to prioritise use cases that present genuine ethical challenges (Challenge 1). Further, data analytics projects evolve dynamically over time. Capturing these changes by team members on a recurring, manual basis helps make an ethical assessment based on the current status quo with the current parameters of the project (Challenge 2). Lastly, employees may require guidance on potential ethical risks in their projects due to a lack of expertise in ethics. These indications can be generated automatically based on certain parameters. The discussion about these risks or strategies should be conducted in person, adapting to the specific context of the project. This ensures that employees are sensitised to continually question the ethical implications of their projects, ensuring that the final ethical assessment always involves human oversight (Challenge 3).

4 Merck's Digital Ethics Check (MDEC)

Based on the preliminary analysis in the previous sections, where we investigated the specific challenges organisations may encounter when integrating ethical principles into the workflow of a data analytics department, and considering

ACE's unique circumstances, a semi-automated approach has emerged as the most appropriate strategy. This approach balances technical capabilities with human oversight, addressing the organisation's specific needs effectively. The strategies implemented during the establishment of the MDEC are intended to act as instructive precedents, providing a framework for other organisations to empower their data analytics teams in identifying and navigating the complexities of digital ethics.

In what follows, we will outline the design of the MDEC as a tool built to synchronise with both Merck's CoDE and the everyday ACE workflow, offering user-friendly digital ethics checkpoints throughout the distinct phases of a typical use case lifecycle. We also provide an overview of its basic functions. The MDEC tool was developed through a collaborative effort involving Merck's Digital Ethics Office, the ACE management and developer teams. This approach ensured that the development process benefited from Merck's expertise in digital ethics and that end-users were involved from the beginning, fostering a collaborative environment, thus ensuring their support, acceptance, and the practical applicability of the outcomes.

4.1 The basic function of the MDEC

The MDEC has been incorporated into the data integration platform as an inherent component of a use case page, ensuring that it becomes a crucial part of the daily workspace and workflow without impeding the work of employees. A fundamental goal of developing the MDEC was to minimise operational burden and bureaucratic overhead for employees. In such an unfavourable scenario, the employees would need to deviate from their usual project management workflow to conduct the MDEC. The integration of the MDEC follows a two-phase methodology, as illustrated in Fig. 1. Each use case is tested twice with the MDEC: initially during the proof-of-value stage, and later during the industrialisation stage to confirm the ethical readiness for deployment.

In what follows, we will explain the MDEC's mechanisms, encapsulating aspects from the initial semi-automated categorisation and risk scoring phases to the final stages involving mitigation strategies and competent stakeholder engagement, all of which enforce the adherence to digital ethics within the realm of data analytics, responding to the challenges outlined above.

The initial stage of the MDEC begins with a semi-automated categorisation of the use case. This process culminates in the computation of an overall risk score, a methodology previously discussed by the AI Ethics Impact Group (2020) or VDE (2022). Based on this score, the MDEC prompts the use case team to engage in a thorough reflective process to identify and address potential critical issues in phase 2 of the MDEC. This engagement is supplemented

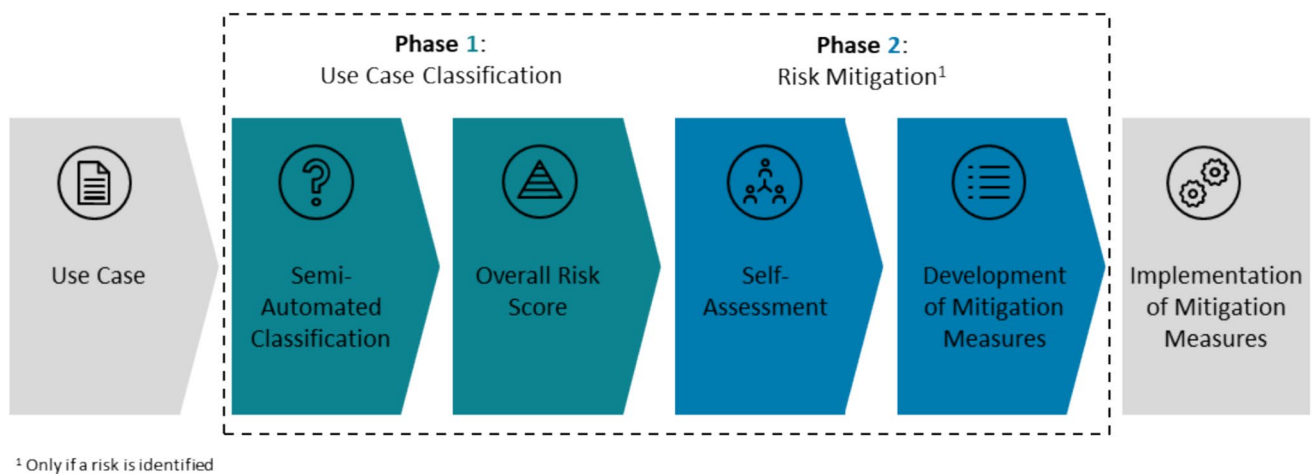


Fig. 1 An overview of Merck's Digital Ethics Check

by an organised self-assessment procedure. Depending on the insights garnered from this process, the product owner may be prompted to formulate strategic measures to mitigate any identified ethical risks. Throughout this stage, further guidance and support are made available to the team through review sessions, which are effectively supported by Merck's Digital Ethics Office.

4.2 The semi-automated classification function of the MDEC

The key component of the semi-automated classification is a questionnaire consisting of a series of multiple-choice questions. These questions span over four critical categories, namely, business characteristics, stakeholders, data, and analysis. The responses to these questions are instrumental to the MDEC's framework, providing the necessary data to calculate the associated risk score of the use case at hand.

This questionnaire extensively explores a spectrum of ethical considerations central to the use case. These multi-dimensional questions establish the foundation of the MDEC's internal structure, ensuring a thorough ethical risk assessment is conducted. Essentially, these considerations hinge on ethically relevant indicators that unite data analytics inquiries with ethical principles encapsulated in the CoDE. The development of the questionnaire in this way is pivotal, primarily because it aligns with Merck's attempts to systematically operationalise digital ethics and adhere to self-imposed standards such as the CoDE. This CoDE not only provides consistent ethical direction within the company but also stands as a declaration of Merck's dedication to uphold specific ethical standards. For organisations lacking an explicitly formulated digital ethics framework, the imperative becomes to either establish such a framework or

to collaboratively define and commit to ethical objectives they aspire to reach.

For the MDEC, the principles of the CoDE have been synthesised into bespoke “ethics markers.” These markers have been carefully calibrated to enhance the evaluation of ethics while promoting the strategic objectives of ACE within the context of data analytics. The underpinning assessment architecture, as depicted in Fig. 2, epitomises a formal structure where these ethics markers are operationalised into measurable “observables.” These observables are intentionally designed to be easily quantifiable by product owners, thereby enabling them to efficiently input relevant data into the MDEC questionnaire. Each step of this process contributes to a common understanding of the ethical implications of the use case, underscored by the principles of the CoDE and facilitated by the structured framework of the MDEC.¹

The MDEC draws extensively from Merck's comprehensive CoDE, integrating ethical considerations in data analytics with a principles-based framework.² To develop a similar tool, organisations need in-depth knowledge of digital ethics tailored to data analytics and must identify crucial parameters for the evaluation of potential risks. Principle-based reasoning behind these parameters is one key to creating ethical markers and clear communication about the tool's objectives and functionality.

To elucidate this methodology, consider the principle of “traceability” as an example. This principle asserts the ethical imperative of consistently monitoring the origin of data,

¹ The other ethics markers of the MDEC can be found in Online Resource 1.

² For further information on the principles, please refer to Becker et al. (2022).

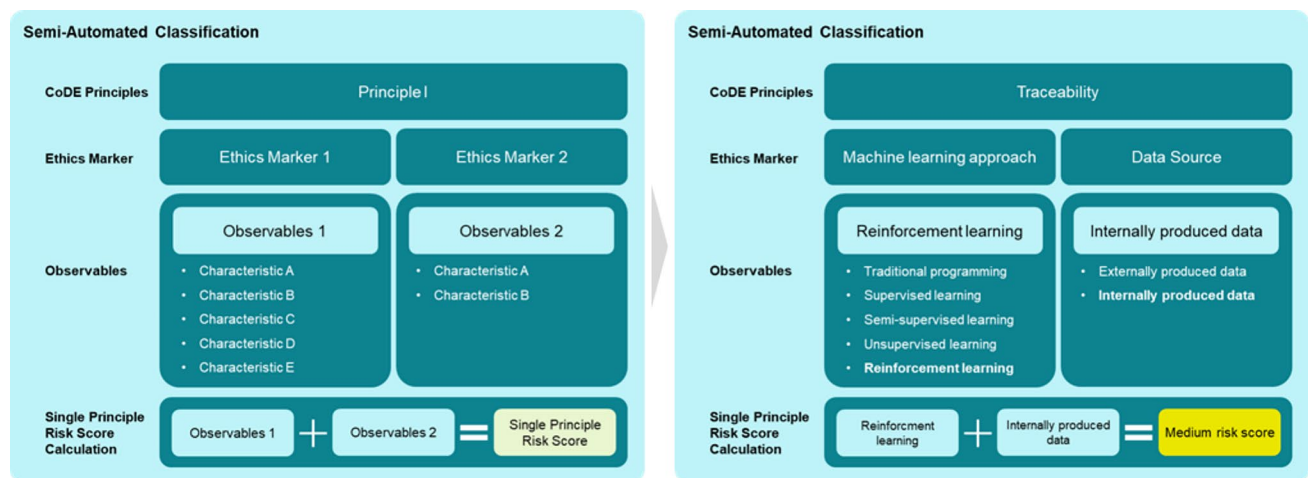


Fig. 2 The semi-automated classification function

tracking its trajectory and interconnections during utilisation, and documenting its subsequent processing. Emphasis is placed on the usage of data sourced from qualified and verified entities, thereby ensuring the data's history is both well-documented and effectively communicable. Within the domain of data analytics projects and specifically within the ACE framework, traceability takes on an important role. It involves record-keeping of the data's journey, an aspect which is linked to the ACE framework's reliance on data integrity. To operationalise traceability, two primary facets are scrutinised: the data's origin and its underlying machine learning approach. Thus, "traceability" has been quantified within the "machine learning approach" and "data source" markers.

Understanding the type of machine learning technique employed in a use case is also important since the complexity involved in tracing data linkages varies across distinct methodologies (High-Level Expert Group on AI 2019). For instance, unsupervised learning methods often pose more challenges in data traceability compared to traditional programming methods, owing to their opaque data connections. The machine learning approach chosen, therefore, acts as a gauge for the significance of the traceability principle. The spectrum of machine learning approaches ranges from "traditional programming" to a variety of machine learning paradigms including "supervised, semi-supervised, unsupervised, and reinforcement learning". Each of these methodologies entails specific considerations in the context of the traceability principle. Accordingly, these models come to serve as "observables"—measurable elements—each incorporated into the MDEC questionnaire.

The traceability principle is not only influenced by the machine learning approach, but also by the data's origin. Data sourced externally is generally deemed more critical, primarily due to the uncertainties inherent to their

provenance. Thus, the ethical considerations associated with traceability intensify when utilising external databases. To assess the degree of traceability, the MDEC differentiates between "externally produced data" and "internally produced data" observables.

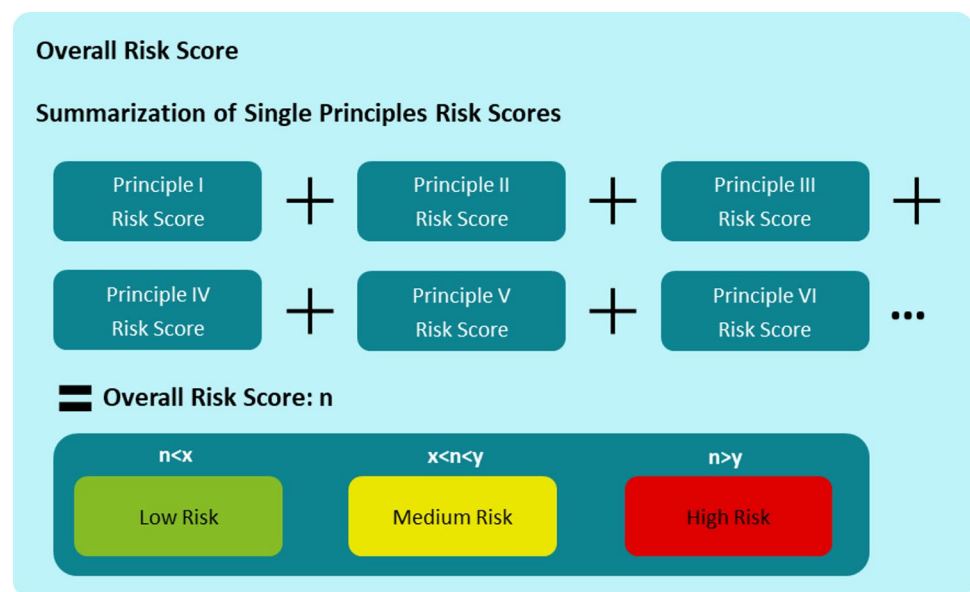
4.3 The calculation of the risk score

Utilising the outlined observables, MDEC calculates a unique risk score for each ethical principle stipulated in the CoDE. This quantifies the ethical risk associated with each specific principle, reflecting a low-, medium-, or high-risk potential concerning adherence to said principle. For example, in the context of traceability, machine learning methodologies like unsupervised and reinforcement learning are considered to harbour high risk due to their extensive data prerequisites and potential for less predictable outcomes, irrespective of the data's source. Conversely, when a supervised learning approach is paired with data from an internal database, the perceived risk is generally low. This pattern continues across the risk evaluations.

It is worth noting that certain markers and observables are relevant concurrently to several principles. For instance, "machine learning approaches" while paramount in evaluating the traceability principle, also bear ethical significance with respect to the principles of "explainability" and "controllability". In the MDEC procedure, individual risk scores for each principle are determined by analysing defined markers and observables.

MDEC aggregates individual risk scores associated with each use case to generate an overall risk score, as illustrated in Fig. 3. Subsequently, the use cases are classified within one of the three risk categories: low, medium, or high. The key determinant in this classification is the establishment of thresholds, which serve as evaluative benchmarks for

Fig. 3 The calculation of the overall risk score



determining the risk level of each use case. For instance, these thresholds specify the score at which a use case qualifies as having a medium risk level.

To customise these thresholds to the unique environmental variables of MDEC, a dual-pronged strategy was implemented. Initially, a preliminary thresholding concept was applied to 35 use cases already in the pipeline within the ACE development process. The use cases were selected to represent a wide variety of use cases developed in ACE. Following this, interviews were conducted with ACE employees to discern the relevance and efficacy of these preliminary results. Insights gleaned from both the evaluation of the use cases and feedback from the employees informed the formation of the final thresholds. The thresholds are defined in the MDEC as follows: 9 principles are used within the MDEC, if more than two of them reach a medium risk level, the whole case is rated as medium risk. If 8 cases are medium and one principle reaches a high risk level, the whole case is rated as high risk. Organisations developing semi-automated ethics assessments should seek to collaborate with stakeholders to fine-tune the sensitivity and thresholds of MDEC-like tools, with user feedback playing a significant role in the development process.

4.4 The self-assessment report and risk mitigation measures

During the second phase of the MDEC, the primary objective revolves around the execution of effective risk mitigation strategies. This phase is initiated on the basis of results derived from phase 1. Here, the self-assessment conducted by the use case team emerges as a key component, playing a substantial role in identifying potential ethical concerns.

Consequently, MDEC generates a detailed report encapsulating the outcome of the assessment.

This report outlines the deployed methodology, summarises the use case, lists the achieved results and scores, and provides a thorough analysis of the findings. A crucial element of the report is the overall score, derived from aggregating individual scores related to valued principles. If the final assessment finds a use case to be low risk, it is assumed that no further steps are required.

However, in instances where use cases are classified as medium or high risk, the second phase takes precedence with a focus on meticulous risk mitigation. In such scenarios, stakeholders are furnished with both an MDEC report and a self-assessment report. This report comprises relevant questions and detailed guidance for self-evaluation. The results of the risk assessment, combined with the current phase of the use case, dictate the need for involving additional stakeholders in the ensuing discussion process. As an exemplary case, during the proof-of-value phase, if a use case is identified as a medium risk, the use case team—composed of the product owner (responsible for the scope of the use case), data steward (responsible for the data), and technical lead (responsible for the implementation)—will engage in discussions guided by insights from the self-assessment report.

In cases where the risk is deemed to be high, it is expected that the Digital Ethics Steering Committee (an internal ACE committee consisting of ACE employees and guests from other departments) is involved in the subsequent discussions to formulate strategies to mitigate the risks identified. This structure extends into the subsequent industrialisation phase. In high-risk scenarios, both the Digital Ethics Steering Committee and a representative from Merck's Digital Ethics

Office are required to participate in discussions guided by the self-assessment report.

Notably, the CoDE principles form the foundational layer of the self-assessment. These principles aid in formulating specific questions aimed at fostering in-depth introspection within the use case teams about their projects from an ethical perspective, consistently aligned with the CoDE principles. For instance, with regard to the traceability principle, the team is encouraged to consider whether data sources are indeed “qualified and verified?” This reflects Merck’s commitment to employing only high-quality, traceable data. Further, supplemental guidance, rooted in both the principle descriptions and supplementary literature research, is provided to support the use case teams throughout the self-assessment process.

This additional guidance provides practical advice, such as considering data transformation processes or simplifying technical complexities, which effectively enhances traceability within the use case. These distilled techniques contribute to a structural and effective ethical assessment, thereby further refining the MDEC approach within the ACE workflow.

An approach like the MDEC, which encourages data analytics teams to discuss specific use cases from an ethical perspective and prioritises human judgement over automated systems in the ethical assessment of use cases, fosters a culture where employees naturally scrutinise their work for ethical considerations. To accomplish this, relying on an established ethics structure, such as a Digital Ethics Office, while potentially favourable for complex organisations such as Merck, may not be necessary for other organisations. Especially for small companies, training employees in digital ethics may sufficiently enhance the quality of ethical discussions. Yet, those employing category-based risk systems should scale their approach accordingly. High-risk cases merit heightened caution, potentially involving extra supervisors for oversight.

5 How the MDEC enables ethical decision-making in data analytics

5.1 Responding to the three challenges

Following our exploration of the MDEC in the previous section, our focus now shifts to its role within the larger dialogue on integrating ethics into data analytics or operationalising digital ethics. We will examine how the MDEC addresses the challenges identified in this area, as outlined in Sect. 2 and consider the potential impact of our findings in further practical contexts. Specifically, we will discuss how the principles and strategies used in the MDEC could potentially be generalised or adapted to other contexts in the field of data analytics. In doing so, we aim to connect

the practicalities of using MDEC with wider conversations concerning ethics in data analytics, providing a grounded understanding of this critical subject.

By responding to the first challenge concerning the high volume of use cases, MDEC integrates a strategic categorisation mechanism. This structure systematically classifies use cases into low-, medium-, or high-risk segments. High-risk use cases necessitate a thorough ethical assessment involving a broader range of stakeholders. Medium-risk scenarios typically involve the internal use case team, while low-risk cases are presumed to require no additional scrutiny, thereby optimising resource allocation and efficiency in managing ethical considerations within a high-volume work environment. To ensure operational scalability, the team deliberately adopted a thoughtful risk-taking approach, thereby avoiding not only carelessness but also excessive caution. MDEC’s sifting mechanism and the resulting stratified review process embody an effective, scalable solution and highlight the potential for wider applicability. In data analytics operating environments characterised by large numbers of projects and iterative workflows, a tool akin to MDEC is therefore invaluable to filter use cases according to the complexity of ethical considerations improving not only the overall efficiency but also enabling a full integration of ethical evaluations into the ordinary work processes without imposing burdensome administrative tasks on the employees.

Concerning the second challenge—the complexities inherent in dynamic project management—the MDEC is designed to accommodate the agile nature of data analytics workflows. It tackles the challenge by adopting a flexible approach that evolves alongside the use case’s lifecycle. By introducing two pivotal checkpoints—one at the transition to the proof-of-value development phase and another before entering the industrialisation phase—MDEC ensures that ethical considerations are constantly being reviewed and assessed. These checkpoints serve as opportunities for reflection and reassessment, enabling the team to identify and address new ethical issues that may arise due to changes in project objectives, methodology, or data sources. It is important to note that the checkpoints are mandatory and that the teams are unable to progress with their project without performing the MDEC. This dynamic and versatile approach ensures that ethical evaluations align concurrently with the project’s progression, offering a responsive and robust solution to managing ethics in an iterative project environment.

In terms of broader applicability beyond MDEC and ACE, the approach of integrating ethics checkpoints throughout a project’s lifecycle could be beneficial in many data analytics settings. As most data analytics projects are characterised by changing strategies and iterative processes, having designated moments for ethical review could significantly enhance the systematic integration of ethical

considerations. Furthermore, this checkpoint-based strategy could be tailored to accommodate varying project dynamics within different organisations, providing a model that could adapt to different scales of operation and unique project cycles. By offering a mechanism that evolves and adapts in response to project development, MDEC’s approach to addressing project dynamics could significantly contribute to the embedding, and reinforcement of ethics within the fluid landscape of data analytics.

Moving on to the third challenge, namely the marked lack of in-depth ethical expertise, MDEC offers an innovative solution by constructing an ethical infrastructure (Floridi 2012) within the use case development lifecycle. Although ACE teams are typically very heterogenous, encompassing both technical specialists and generalists, the realm of data and AI ethics requires unique expertise, which is typically absent within these teams. MDEC addresses this lack of expertise by establishing a robust framework that fosters a collective consciousness for ethical considerations. By actively encouraging the team to critically interrogate use cases from an ethical perspective and stimulating open discussions about ethically pertinent matters, MDEC instigates a broader awareness of ethical considerations and encourages an ongoing dialogue about the ethical aspects of data analytics work.

5.2 Implications for practitioners and researchers

Broadly speaking, the concept of instilling an ethical infrastructure within data analytics operations holds significant potential for wider application beyond MDEC and ACE, potentially benefitting other organisations embedding similar mechanisms. By fostering widespread ethical awareness and providing a model for structured ethical assessment, organisations can integrate ethical considerations throughout their workflows in a manner that is both accessible and uncomplicated, effectively bridging the gap between ethical theory and practice. As such, MDEC’s strategic approach to cultivating ethical expertise offers a valuable impetus for broader integration of ethical considerations within the realm of data analytics or AI applications that pose similar

challenges, irrespective of existing expertise levels within the teams. This can significantly influence decision-making processes, ensure accountability, and ultimately anchor an ethical mindset within the data analytics workforce.

In parallel, the methodology adopted to develop a framework such as MDEC can function not merely as a resource for practitioners but also as a foundation for further academic inquiry into semi-automated approaches to risk assessment. Specifically, the process by which broader principles are translated into concrete, ethics-oriented tools ought to be refined and subjected to scholarly scrutiny. Such efforts benefit not only AI practitioners but also those who work with broader questions of translational ethics. The key take-aways of our work are summarised in Table 1.

5.3 Further development

Prior to its implementation, the MDEC underwent extensive testing with a range of use cases and user profiles within ACE’s data analytics unit. In its final iteration, the tool’s outcomes proved largely consistent with both the use case owners’ initial intuitions and the Bioethics and Digital Ethics Office’s comprehensive ethical assessments. Its ongoing deployment, however, requires continued monitoring to evaluate accuracy and recalibrate parameters as needed. After the first few months of implementation, the MDEC’s outputs are currently under review to pinpoint false positives and negatives and to identify recurring themes among medium- and high-risk use cases. These findings will serve to iteratively refine and enhance the tool.

Moreover, the MDEC in its present form has been tailored specifically to the operational needs and process-related constraints emanating from the data analytics environment within ACE. Nonetheless, other departments facing similar challenges—particularly those related to dynamic project requirements and high throughput—have also begun to employ AI tools extensively. To accommodate these differing contexts, efforts are currently underway to develop adapted versions of the MDEC, calibrated to each department’s distinct needs. For instance, certain markers may be redefined or accorded varying levels of importance; in

Table 1 Key take-aways of our work

Key take-aways
The development of an ethics evaluation tool is most effective if undertaken in close collaboration with the department that will use the tool
The implementation of the tool is likely to yield broad acceptance if integrated into the existing use case process with a view to avoiding the creation of additional structures for employees
The utilisation of a semi-automated approach enables a harmonisation between an autonomised evaluation and human involvement
Effective fine-tuning, for example through both ex-ante and ex-post implementation evaluation, is important to ensure the tool’s functionality
The involvement of an external ethics panel or other forms of independent review can be beneficial in challenging the development of an ethics evaluation tool

human resources, for example, greater emphasis is placed on ensuring that data subjects are informed about their data usage (a newly introduced marker), while the data-type category has been refined to include more specific classifications such as biometric information (an adaptation of an existing marker). Likewise, for Digital Health applications, additional attention must be devoted to the intended target group, recognising that solutions addressing patients, caregivers, or healthcare personnel entail heightened ethical considerations. These initiatives are ongoing, and further details will be reported in separate publications.

6 Conclusion

The present paper exemplifies the advantages of a structured, semi-automated integration of digital ethics into data analytics, as illustrated by the MDEC, in addressing unique challenges that hinder individual consultation while also serving other purposes such as enhancing ethical competency. It provides a practical impetus to guide organisations intending to embed ethical considerations within their data analytics operations. Illustrative insights into the MDEC's principles, mechanisms, and strategies, demonstrated through its dual-phase approach, risk categorisation, and continuous evaluation, contribute to the understanding of the challenges and opportunities of operationalising digital ethics. Thus, this examination offers valuable insights adaptable across a diverse range of organisational environments. The advantages and limitations of our approach are summarised in Table 2.

For the further development of the MDEC, it is essential to maintain a consistent and rigorous approach to the monitoring and analysis of its results. This necessitates a thorough examination of the extent to which the MDEC is generating false positive or false negative outcomes, as well as an investigation into the underlying causes of such

discrepancies. Additionally, it is crucial to recognise the MDEC as an evolving system that may require adaptation over time to ensure its continued efficacy. In order to guarantee this, the management of ACE and the Digital Ethics Office will collaborate closely in order to further develop the MDEC.

However, this work extends beyond a mere case study of the MDEC. It opens the door to a deeper understanding of the complexities involved in integrating ethics within the continually evolving field of data analytics. By discussing major challenges such as managing a vast array of use cases, navigating complex project management structures, and addressing the expertise gap in ethics within data analytics teams, we contribute to the broader discourse at the intersection of ethics and digital technology. Therefore, this paper presents an exemplary view of meeting real-world challenges and devising strategic solutions—highlighting critical areas for discussion within the nexus of technology, work, and ethics, many of which are applicable more broadly to data analytics and the use of AI in corporate settings.

Thus, our study makes a practical contribution to demonstrating how bridging the research gap in the operationalisation of digital ethics is feasible, with a particular focus on the realm of data analytics. It depicts a real-life scenario that captures the dynamics of ethical deliberations within data-centric corporate settings, thereby paving the way for practical discourses that emphasise the need for ongoing research to address the practical applications of feedback-oriented ethics mechanisms in data-intensive sectors.

In summary, this work attests to the critical relevance of ethics in the context of continuous technological advancement. It provides an exemplary case that illuminates the complexities faced by those seeking to navigate the complex task of integrating digital ethics into data-driven decision-making processes. The insights derived from this investigation encourage further academic exploration into the connection between data analytics and ethical principles.

Table 2 Advantages and limitations

Advantages	Limitations
A semi-automated approach provides automated filtering for a wide range of use cases, while ensuring the necessary human involvement to guarantee human control	The automation of processes creates the possibility of missing critical cases. Therefore, it is essential that employees receive additional ethics training, enabling them to challenge projects independently when necessary
The incorporation of a tool within the standard use case lifecycle ensures a seamless integration and safeguards employees against the risk of bureaucratic exhaustion	The use of a questionnaire for identifying ethical issues is constrained by the format of interrogation and the number of questions. Therefore, additional guidance and employee-led discussions are encouraged within the subsequent self-assessment
Empowering employees through the implementation of an ethical self-assessment and training promotes a bottom-up approach to creating an “ethics culture”	Employees may feel overwhelmed when confronted with ethically complex use cases or disengaged if their own judgements are seen as being paternalistically undermined. Therefore, it is important that critical cases are addressed through individual collaboration with the Digital Ethics Office

Consequently, the study contributes to a critical dialogue aimed at fostering ethical integrity in the data-dense era.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00146-025-02621-2>.

Author contributions The authors contributed equally.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data availability No data sets were generated or analysed during the current study.

Declarations

Conflict of interest All authors declare that they have no conflict of interest to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ada Lovelace Institute (2022) Algorithmic impact assessment: a case study in healthcare. Ada Lovelace Institute. <https://www.adalovelaceinstitute.org/wp-content/uploads/2022/02/Algorithmic-impact-assessment-a-case-study-in-healthcare.pdf>. Accessed 12 Jun 2024
- Adams J (2024) Introducing the ethical-epistemic matrix: a principle-based tool for evaluating artificial intelligence in medicine. *AI Ethics*. <https://doi.org/10.1007/s43681-024-00597-1>
- AI Ethics Impact Group (2020) From principles to practice: An interdisciplinary framework to operationalise AI ethics. AI Ethics Impact Group. <https://www.ai-ethics-impact.org/en>. Accessed 08 Oct 2023
- Aiken C (2021) Classifying AI systems CSET data brief. CSET. <https://cset.georgetown.edu/publication/classifying-ai-systems/>. Accessed 12 Jun 2024
- Aizaz F, Khan S, Khan JA, Rehman I, Akhuzada A (2021) An empirical investigation of factors causing scope creep in agile global software development context: a conceptual model for project managers. *IEEE*. <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9499039>. Accessed 12 Jun 2024
- Amugongo LM, Kriebitz A, Boch A, Lütge C (2023) Operationalising AI ethics through the agile software development lifecycle: a case study of AI-enabled mobile health applications. *AI Ethics*. <https://doi.org/10.1007/s43681-023-00331-3>
- Becker SJ, Nemat AT, Lucas S, Heinitz RM, Klevesath M, Charton J (2022) A Code of Digital Ethics: Laying the foundation for digital ethics in a science and technology company. *AI Soc*. <https://doi.org/10.1007/s00146-021-01376-w>
- Bleher H, Braun M (2023) Reflections on putting AI ethics into practice: how three AI ethics approaches conceptualize theory and practice. *Sci Eng Ethics*. <https://doi.org/10.1007/s11948-023-00443-3>
- Brown S, Davidovic J, Hasan A (2021) The algorithm audit: scoring the algorithms that score us. *Big Data Soc*. <https://doi.org/10.1177/2053951720983865>
- Char DS, Abràmoff MD, Feudtner C (2020) Identifying ethical considerations for machine learning healthcare applications. *Am J Bioeth* 20(11):7–17. <https://doi.org/10.1080/15265161.2020.1819469>
- Chmielinski KS, Newman S, Taylor M, Joseph J, Thomas K, Yurkofsky J, Qiu YC (2020) The dataset nutrition label (2nd gen): leveraging context to mitigate harms in artificial intelligence. *NeurIPS 2020 Workshop on Dataset Curation and Security*. http://securedata.lol/camera_ready/26.pdf
- Cobbe J, Lee MSA, Singh J (2021) Reviewable automated decision-making: a framework for accountable algorithmic systems. In: *FaccT'21: proceedings of the 2021 ACM conference on fairness, accountability, and transparency*. <https://doi.org/10.1145/3442188.3445921>
- Corrêa NK, Santos JW, Galvão C, Pasetti M, Schiavon D, Naqvi F, Hossain R, Oliveira ND (2024) Crossing the principle–practice gap in AI ethics with ethical problem-solving. *AI Ethics*. <https://doi.org/10.1007/s43681-024-00469-8>
- Danish Design Center (2021) The digital ethics compass. Danish Design Center. <https://ddc.dk/tools/toolkit-the-digital-ethics-compass>. Accessed 12 Jun 2024
- Danks D, London AJ (2017) Algorithmic bias in autonomous systems. In: *Twenty-sixth international joint conference on artificial intelligence*. Algorithmic bias in autonomous systems. <https://www.ijcai.org/Proceedings/2017/0654.pdf>
- d'Aquin M, Troullinou P, O'Connor NE, Cullen A, Faller G, Holden L (2018) Towards an “ethics by design” methodology for AI research projects. In: *Proceedings of the 2018 AAAI/ACM conference on AI, ethics, and society*. pp 54–59. <https://doi.org/10.1145/3278721.3278765>
- De Bie T, De Raedt L, Hernández-Orallo J, Hoos HH, Smyth P, Williams CK (2022) Automating data science. *Commun ACM* 65(3):76–87. <https://doi.org/10.1145/3495256>
- Deon (2018) An ethics checklist for data scientists. Deon. <https://deon.drivendata.org/>. Accessed 28 Apr 2025
- Deutscher Ethikrat (2017) Big Data und Gesundheit—Datensouveränität als informationelle Freiheitsgestaltung. Deutscher Ethikrat. <https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-big-data-und-gesundheit.pdf>. Accessed 12 Jun 2024
- Díaz de la Cruz C, Fernández Fernández JL, Villegas-Galaviz C (2025) Model of ethical analysis of digital technologies: towards true digital humanism. *AI Soc*. <https://doi.org/10.1007/s00146-025-02330-w>
- Epstein Z, Payne BH, Shen JH, Hong CJ, Felbo B, Dubey A, Groh M, Obradovich N, Cebrian M, Rahwan I (2018) TuringBox: an experimental platform for the evaluation of AI systems. In: *Proceedings of the twenty-seventh international joint conference on artificial intelligence*. <https://doi.org/10.24963/ijcai.2018/851>
- Fjeld J, Achten N, Hilligoss H, Nagy A, Srikumar M (2020) Principled artificial intelligence: mapping consensus in ethical and rights-based approaches to principles for AI. *Berkman Klein Center Research Publication*, No. 2020–2021. <https://doi.org/10.2139/ssrn.3518482>
- Floridi L (2012) Distributed morality in an information society. *Sci Eng Ethics* 19:727–743. <https://doi.org/10.1007/s11948-012-9413-4>

- Floridi L (2019) Translating principles into practices of digital ethics: five risks of being unethical. *Philos Technol* 32:185–193. <https://doi.org/10.1007/s13347-019-00354-x>
- Floridi L, Sanders JW (2004) On the morality of artificial agents. SSRN. <https://ssrn.com/abstract=3848388>
- Floridi L, Cowls J, Beltrame M, Chatila R, Chazerand P, Dignum V, Luetge C, Madelin R, Pagallo U, Rossi F, Schafer B, Valcke P, Vayena E (2018) AI4People - An ethical framework for a good AI society: opportunities, risks, principles, and recommendations. *Minds Mach* 28:689–707. <https://doi.org/10.1007/s11023-018-9482-5>
- Floridi L, Holweg M, Taddeo M, Silva JA, Mökander J, Wen Y (2022) capAI—a procedure for conducting conformity assessment of AI systems in line with the EU artificial intelligence act. SSRN. <https://ssrn.com/abstract=4064091>
- ForHumanity (2021) Independent audit of AI systems. ForHumanity. <https://forhumanity.center/independent-audit-of-ai-systems>. Accessed 09 Sept 2022
- Georgieva I, Lazo C, Timan T, Van Veenstra AF (2022) From AI ethics principles to data science practice: a reflection and a gap analysis based on recent frameworks and practical experience. *AI Ethics* 2(4):697–711. <https://doi.org/10.1007/s43681-021-00127-3>
- Giovanola B, Tiribelli S (2023) Beyond bias and discrimination: redefining the AI ethics principle of fairness in healthcare machine-learning algorithms. *AI Soc* 38(2):549–563. <https://doi.org/10.1007/s00146-022-01455-6>
- Gotterbarn D, Krepis D (2021) Being a data professional: give voice to value in a data driven society. *AI Ethics* 1:195–203. <https://doi.org/10.1007/s43681-020-00027-y>
- Grellette M (2022) Tech ethics through trust auditing. *Sci Eng Ethics*. <https://doi.org/10.1007/s11948-022-00379-0>
- Hagendorff T (2020) The ethics of AI ethics: an evaluation of guidelines. *Minds Mach* 30:99–120. <https://doi.org/10.1007/s11023-020-09517-8>
- Hickok M (2020) Lessons learned from AI ethics principles for future actions. *AI Ethics* 1:41–47. <https://doi.org/10.1007/s43681-020-00008-1>
- High-Level Expert Group on AI (2019) Ethics guidelines for trustworthy AI. High-Level Expert Group on AI. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>. Accessed 12 Jun 2024
- Hirsch D, Bartely T, Chandrasekaran A, Norris D, Parthasarathy S, Turner PN (2021) Business data ethics: emerging trends in the governance of advanced analytics and AI. Final report. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3828239
- Jobin A, Ienca M, Vayena E (2019) The global landscape of AI ethics guidelines. *Nat Mach Intell* 1:389–399. <https://doi.org/10.1038/s42256-019-0088-2>
- Kazim E, Mendes Thame Denny D, Koshiyama A (2021) AI auditing and impact assessment: according to the UK information commissioner's office. *AI Ethics* 1:301–310. <https://doi.org/10.1007/s43681-021-00039-2>
- Keller SA, Shipp SS, Schroeder AD, Korkmaz G (2020) Doing data science: a framework and case study. *Harv Data Sci Rev*. <https://doi.org/10.1162/99608f92.2d837f5>
- Kelley S (2022) Employee perceptions of the effective adoption of AI principles. *J Bus Ethics* 178:871–893. <https://doi.org/10.1007/s10551-022-05051-y>
- Kraus S, Durst S, Ferreira JJ, Veiga P, Kailer N, Weinmann A (2021) Digital transformation in business and management research: an overview of the current status quo. *IJIM*. <https://doi.org/10.1016/j.ijinfomgt.2021.102466>
- Krijger J, Thuis T, de Ruiter M, Ligthart E, Broekman I (2023) The AI ethics maturity model: a holistic approach to advancing ethical data science in organizations. *AI Ethics* 3(2):355–367. <https://doi.org/10.1007/s43681-022-00228-7>
- Kroll JA (2018) Data science data governance. NIST. https://www.nist.gov/system/files/documents/2019/06/10/nist-ai-rfi-kroll-data_science_data_governance.pdf. Accessed 12 Jun 2024
- Law PM, Malik S, Du F, Sinha M (2020) Designing tools for semi-automated detection of machine learning biases: an interview study. arXiv preprint. <https://arxiv.org/pdf/2003.07680>
- Leonelli S (2016) Locating ethics in data science: responsibility and accountability in global and distributed knowledge production systems. *Philos Trans R Soc Lond A Math Phys Eng Sci* 374:20160122. <https://doi.org/10.1098/rsta.2016.0122>
- Martin K (2022) Ethics of data and analytics concepts and cases. CRC Press, Taylor & Francis Group, Boca Raton
- McLennan S, Fiske A, Tigard D, Müller R, Haddadin S, Buyx A (2022) Embedded ethics: a proposal for integrating ethics into the development of medical AI. *BMC Med Ethics* 23:6. <https://doi.org/10.1186/s12910-022-00746-3>
- Meier LJ, Hein A, Diepold K, Buyx A (2022) Algorithms for ethical decision-making in the clinic: a proof of concept. *Am J Bioeth* 22(7):4–20. <https://doi.org/10.1080/15265161.2022.2040647>
- Merck KGaA (2021) Merck announces formation of Merck Digital Ethics Advisory Panel. <https://www.merckgroup.com/press-releases/2021/jan/en/Digital-Ethics-Advisory-Panel-EN.pdf>. Accessed 12 June 2024
- Mittelstadt BD (2019) Principles alone cannot guarantee ethical AI. *Nat Mach Intell* 1:501–507. <https://doi.org/10.1038/s42256-019-0114-4>
- Mittelstadt BD, Floridi L (2016) The ethics of big data: current and foreseeable issues in biomedical contexts. *Sci Eng Ethics* 22(2):303–341. <https://doi.org/10.1007/s11948-015-9652-2>
- Mökander J, Floridi L (2021) Ethics-based auditing to develop trustworthy AI. *Minds Mach* 31:23–327. <https://doi.org/10.1007/s11023-021-09557-8>
- Mökander J, Floridi L (2022) Operationalising AI governance through ethics-based auditing: an industry case study. *AI Ethics*. <https://doi.org/10.1007/s43681-022-00171-7>
- Mökander J, Morley J, Taddeo M, Floridi L (2021) Ethics-based auditing of automated decision-making systems: nature, scope, and limitations. *Sci Eng Ethics*. <https://doi.org/10.1007/s11948-021-00319-4>
- Mökander J, Axente M, Casolari F, Floridi L (2022) Conformity assessments and post-market monitoring: a guide to the role of auditing in the proposed European AI regulation. *Minds Mach* 32(2):241–268. <https://doi.org/10.1007/s11023-021-09577-4>
- Mökander J, Sheth M, Watson DS, Floridi L (2023) The switch, the ladder, and the matrix: models for classifying AI systems. *Minds Mach* 33:221–248. <https://doi.org/10.1007/s11023-022-09620-y>
- Morley J, Floridi L, Kinsey L, Elhalal A (2020) From what to how: an initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Sci Eng Ethics* 26:2141–2168. <https://doi.org/10.1007/s11948-019-00165-5>
- Mühlhoff R (2021) Predictive privacy: towards an applied ethics of data analytics. *Ethics Inf Technol* 23:657–690. <https://doi.org/10.1007/s10676-021-09606-x>
- Nemat AT, Becker SJ, Lucas S, Thomas S, Gadea I, Charton J (2023) The Principle-at-Risk Analysis (PaRA): Operationalising Digital Ethics by Bridging Principles and Operations of a Digital Ethics Advisory Panel. *Minds Mach* 33:737–760. <https://doi.org/10.1007/s11023-023-09654-w>
- Novak R, Pavlicek A (2021) Data experts as the balancing power of big data ethics. *Information* 12(3):97. <https://doi.org/10.3390/info12030097>
- Omidyar Network (2018) New toolkit shows companies how to anticipate, prevent bad actors from using tech in harmful ways. Omidyar Network. <https://medium.com/omidyar-network/intro>

- ducing-the-worlds-first-ethical-operating-system-7acc4abc2bfa. Accessed 12 Jun 2024
- Open Data Institute (2021) The data ethics canvas. Open Data Institute. <https://theodi.org/article/the-data-ethics-canvas-2021/>. Accessed 12 Jun 2024
- Oschmann S (2018) Curious about the future. *Angewandte Chemie International Edition* 57(16):4108–4109. <https://doi.org/10.1002/anie.201711265>
- Passi S, Sengers P (2020) Making data science systems work. *Big Data Soc* 7(2):1–13. <https://doi.org/10.1177/2053951720939605>
- Poretschkin M, Schmitz A, Akila M, Adilova L, Becker D, Cremers AB, Hecker D, Houben S, Mock M, Rosenzweig J, Sicking J, Schulz E, Voss A, Wrobe, S (2021) Leitfaden zur Gestaltung vertrauenswürdiger künstlicher Intelligenz. Fraunhofer IAIS. www.iais.fraunhofer.de/ki-pruefkatolog. Accessed 12 Jun 2024
- Priyono A, Moin A, Putri VNAO (2020) Identifying digital transformation paths in the business model of SMEs during the COVID-19 pandemic. *JOItmC*. <https://doi.org/10.3390/joitmc6040104>
- Rochel J, Évéquoz F (2021) Getting into the engine room: a blueprint to investigate the shadowy steps of AI ethics. *AI Soc* 36:609–622. <https://doi.org/10.1007/s00146-020-01069-w>
- Rudschies C, Schneider I, Simon J (2021) Value pluralism in the AI Ethics debate—different actors, different priorities. *IRIE*. <https://doi.org/10.29173/irie419>
- Ryan M, De Roo N, Wang H, Blok V, Atik C (2024) AI through the looking glass: an empirical study of structural social and ethical challenges in AI. *AI Soc*. <https://doi.org/10.1007/s00146-024-02146-0>
- Saleiro P, Kuester B, Hinkson L, London J, Stevens A, Anisfeld A, Rodolfa KT, Ghani R (2018) Aequitas: a bias and fairness audit toolkit. *arXiv preprint*. <https://arxiv.org/abs/1811.05577>
- Schiff D, Rakova B, Ayesh A, Fanti A, Lennon M (2021) Explaining the principles to practice gap in AI. *IEEE* 40(2):1–94. <https://doi.org/10.1109/MTS.2021.3056286>
- Schuett J, Reuel A, Carlier A (2023) How to design an AI ethics board. *AI Ethics*. <https://doi.org/10.1007/s43681-023-00409-y>
- Selbst AD, Boyd D, Friedler S, Venkatasubramanian S, Vertesi J (2019) Fairness and abstraction in sociotechnical systems. In: *FAT* '19: proceedings of the conference on fairness, accountability, and transparency*. <https://doi.org/10.1145/3287560.3287598>
- Sugarman J, Shivakumar S, Rook M, Loring JF, Rehmann-Sutter C, Taupitz J, Reinhard-Rupp J, Hildemann S (2018) Ethical considerations in the manufacture, sale, and distribution of genome editing technologies. *Am J Bioeth* 18(8):3–6. <https://doi.org/10.1080/15265161.2018.1489653>
- UK Statistics Authority (2022) Guidelines on using the ethics self-assessment process. UK Statistics Authority. https://uksa.statisticsauthority.gov.uk/wp-content/uploads/2021/04/2021_Self-assessment_guidance_V2.3.pdf. Accessed 12 Jun 2024
- Umbrello S (2020) Beneficial artificial intelligence coordination by means of a value sensitive design approach. *Big Data Cogn Comput*. <https://doi.org/10.3390/bdcc3010005>
- Umbrello S, Gambelin O (2021) A value sensitive design toolkit for agile project management. *Academia Letters*. <https://api.semanticscholar.org/CorpusID:247472866>
- UN Global Pulse (2019) Risks, harms and benefits assessment. UN Global Pulse. <https://www.unglobalpulse.org/document/risks-harms-and-benefits-assessment-tool/>. Accessed 12 Jun 2024
- VDE (2022) VDE SPEC 90012 V1.0—VCIO based description of systems for AI trustworthiness characterization. VDE. <https://www.vde.com/resource/blob/2242194/a24b13db01773747e6b7bba4ce20ea60/vcio-based-description-of-systems-for-ai-trustworthiness-characterisationvde-spec-90012-v1-0--en--data.pdf>. Accessed 12 Jun 2024
- Wachter S, Mittelstadt B, Russell C (2021) Why fairness cannot be automated: bridging the gap between EU non-discrimination law and AI. *Comput Law Secur Rev*. <https://doi.org/10.1016/j.clsr.2021.105567>
- Wang D, Weisz JD, Muller M, Ram P, Geyer W, Dugan C, Tausczik Y, Samulowitz H, Gray A (2019) Human-AI collaboration in data science. *Proc ACM Hum Comput Interact* 3(CSCW):1–24. <https://doi.org/10.1145/3359313>
- Wirtz J, Kunz WH, Hartley N, Tarbit J (2023) Corporate digital responsibility in service firms and their ecosystems. *J Serv Res* 26(2):173–190. <https://doi.org/10.1177/10946705221130467>
- Zicari RV, Brodersen J, Brusseau J, Dudder B, Eichhorn T, Ivanov T, Kararigas G, Kringen P, McCullough M, Möslin F, Mushtag N, Roig G, Stürtz N, Tolle K, Tithi JJ, Van Halem I, Westerlund M (2021) Z-Inspection®: a process to assess trustworthy AI. *IEEE Trans Technol Soc* 2(2):83–97. <https://doi.org/10.1109/TTS.2021.3066209>
- Zorio S (2021) How a paper by three Oxford academics influenced AWS bias and explainability software. *Amazon Science*. <https://www.amazon.science/latest-news/how-a-paper-by-three-oxford-academics-influenced-aws-bias-and-explainability-software>. Accessed 28 Apr 2025
- Zuber N, Gogoll J, Kacianka S, Pretschner A, Nida-Rümeling J (2022) Empowered and embedded: ethics and agile processes. *Humanit Soc Sci Commun* 9:191. <https://doi.org/10.1057/s41599-022-01206-4>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.