The Principle-at-Risk Analysis (PaRA): Operationalising Digital Ethics by Bridging Principles and Operations of a Digital Ethics Advisory Panel

André T. Nemat[1] · Sarah J. Becker[1] · Simon Lucas[2] · Sean Thomas[3] · Isabel Gadea[1] · Jean Enno Charton[2]

Abstract

Recent attempts to develop and apply digital ethics principles to address the challenges of the digital transformation leave organisations with an operationalisation gap. To successfully implement such guidance, they must find ways to translate high-level ethics frameworks into practical methods and tools that match their specific workflows and needs. Here, we describe the development of a standardised risk assessment tool, the Principle-at-Risk Analysis (PaRA), as a means to close this operationalisation gap for a key level of the ethics infrastructure at many organisations – the work of an interdisciplinary ethics panel. The PaRA tool serves to guide and harmonise the work of the Digital Ethics Advisory Panel at the multinational science and technology company Merck KGaA in alignment with the principles outlined in the company's Code of Digital Ethics. We examine how such a tool can be used as part of a multifaceted approach to operationalise high-level principles at an organisational level and provide general requirements for its implementation. We showcase its application in an example case dealing with the comprehensibility of consent forms in a data-sharing context at Syntropy, a collaborative technology platform for clinical research.

Keywords Digital ethics · Operationalisation of ethics · Ethical guidelines · Ethical panels · Informed consent

✉ Sarah J. Becker
sarah.becker@transforming-healthcare.com

1   University of Witten/Herdecke, Institute for Digital Transformation in Healthcare, Witten, Germany

2   Merck KGaA, Bioethics & Digital Ethics, Corporate Affairs, Darmstadt, Germany

3   EMD Digital Inc., Burlington, USA

1 Introduction

The pace and nature of the ongoing digital transformation present societal and ethical challenges that require practical guidance in various societal contexts, including business settings. Rich and varied troves of data continue to accrue, while the algorithms used to mine this data grow increasingly sophisticated. Rapid innovation in the scope and performance of digital technologies such as big data analytics, mobile and cloud computing, machine learning, artificial intelligence (AI) and complex networking of devices and their users is arguably changing how humans understand themselves, each other and the world around them (Floridi, 2014).

The advancement of digital technologies can impact various stakeholders, whose needs, values and perspectives may differ (Mittelstadt, 2019; Rudschies et al., 2021). In the business world, for example, digital technologies are transforming how companies operate internally and externally, introducing novel ways to create revenue, improve business and transform industry processes (Kraus et al., 2021). The COVID-19 pandemic has further accelerated this trend in recent years (Priyono et al., 2020). However, digital innovation must balance commercial interests, regulatory developments and uncertainties, and ethical and societal expectations to maintain the trust of customers and business partners.

Challenges range from privacy concerns to issues of bias, explainability, responsibility and transparency (Borrego-Díaz & Galán-Páez, 2022; Buijsman, 2022; Mittelstadt & Floridi, 2016; Rochel & Evéquoz, 2021). To ensure that digital technologies benefit society and avoid harm, efforts have been made to identify ethical principles that should be safeguarded during the development, deployment and use of digital innovations (Floridi et al., 2018; Jobin et al., 2019; Fjeld et al., 2020; Hagendorff, 2020), and many organisations have adopted ethics guidelines based on these principles. Such guidelines offer a proactive solution for addressing digital ethics challenges when regulatory frameworks are slow to adapt to technical and business developments (European Commission, 2021; Gordon et al., 2021; Mökander and Floridi, 2022a). This can serve to foster trust, especially in industry settings where large amounts of potentially sensitive data are mined – for example in the science, technology, and medical sectors. By 2019, there were at least 84 documents proposing principles or guidelines for AI ethics alone, with one-fifth of them developed by private companies (Jobin et al., 2019).

1.1 The Search for a Practical Approach to Ethical Challenges

Recent academic discourse also, however, increasingly recognises that principle-based ethics guidelines must be translated into practice to become more than mere lip service. This requires operationalising high-level guidance to match an organisation's structure, workflows and unique digital challenges (Floridi, 2019; Mittelstadt, 2019; Morley et al., 2020; Hickok, 2020; Blackman, 2020; Mökander et al., 2022b). Several proposals to tackle this operationalisation have been made in academic literature and practical initiatives. Some approaches focus on external ethics-based auditing and certification frameworks to ensure that digital applications and the organisations that employ them meet ethical standards (Epstein et al., 2018; Saleiro et al., 2018;

AI Ethics Impact Group, 2020; Chmielinski, 2020; Cobbe et al., 2021; ForHumanity, 2021; Poretschkin, 2021; Grellette, 2022; Mökander et al., 2021, 2022a; Zicari et al., 2021; Floridi et al., 2022; Mökander & Floridi, 2021, 2022b). Others involve internal self-assessments and impact assessments to guide individuals and organisations to explore the ethical impacts of their digital endeavours, to ground decisions ethically and encourage appropriate evaluation, reflection and, where necessary, mitigation.

Tools may be applied at different stages in a project's life cycle, i.e., at the business and use-case development stage, in the design phase, during training and test data procurement, or while building, testing, deploying, or monitoring an application (Morley et al., 2020). They also differ in their complexity and specificity and in how they integrate into workflows. Some appear as simple checklists that can be used at multiple stages or be built directly into software (Deon, 2018; Keller et al., 2020; Open Data Institute, 2021). Other tools offer to automatically check code or data for ethically relevant issues such as bias (Epstein et al., 2018; Saleiro et al., 2018; Chmielinski, 2020; Wachter et al., 2021; Zorio, 2021). Yet other approaches involve training, workshops, or exercises for teams (Danish Design Center, 2021; Institute for the Future, 2018), and there are also comprehensive methodologies for assessing risks and identifying possible mitigations (High-Level Expert Group on AI, 2019; UN Global Pulse, 2019; Groves, 2022; UK Statistics Authority, 2022).

1.2 Current Practical Approaches and Their Limitations

Despite such a plenitude of tools, there often remains a substantial gap between their availability and successful implementation in practice. Analyses of this "principle-to-practices" gap (Schiff et al., 2021) have led to complimentary explanations and criticisms. Some address the suitability of the existing tools themselves, describing them as not actionable, "as they offer little help on how to use them in practice" (Morley et al., 2020). Many tools are designed to be used as a one-off and put too much emphasis on diagnosing ethical problems, with too little support on how to address them. Furthermore, not enough work has been done to subject the tools themselves to empirical evaluation (Morley et al., 2021a). The sheer number of available tools can also be problematic, as evaluating such an over-abundance and selecting the most suitable candidates puts a strain on people's limited "time, attention, and cognitive capacity […], leading to search and transaction cost problems" (Schiff et al., 2021).

In addition, there is much diversity among digital endeavours and applications, with various actors involved, and different aspects of society affected. This complexity makes it more challenging to implement ethical principles than in other fields, such as medicine, that are more narrowly focussed and clearly structured (Mittelstadt, 2019; Schiff et al., 2021). Therefore, it is crucial to pay attention to specific organisational structures and practices when attempting to operationalise digital ethics. Failing to consider the most promising approach for each situation can lead to issues such as a lack of clarity regarding which systems and processes ethical principles ought to apply to (Mökander et al., 2022a), confusion around roles and responsibilities in their implementation (Morley et al., 2021b), or uncertainty about how to proceed when principles conflict with each other (Sanderson et al., 2021).

These contributions highlight the importance of clear procedures that yet remain adaptable to specific contexts and insights generated during evaluation. Effective operationalisation needs to strike the right balance between making tools and methods too flexible and vague – and thus vulnerable to being misused for avoidance tactics such as "ethics shopping" or "ethics washing" (Floridi, 2019) – or too strict, and thus unsuitable to the dynamic complexities and specific contexts of digital ethics challenges (Morley et al., 2021a, b).

To achieve such a balance, it has been recommended that future operationalisation efforts should focus on organisational ethics and ethics as a process (Mittelstadt, 2019) and be "continuous, holistic, dialectic, strategic and design-driven" (Mökander & Floridi, 2022b). They should involve several components such as independent multi-disciplinary ethics boards, ethics codes and individual practitioners (Morley et al., 2021a); and include explicit governance models to clarify responsibilities (Georgieva et al., 2022). Any such strategies will have to be multifaceted, selecting and combining different tools and approaches to suit the specific situation as well as the organisation and its component bodies, teams and other structural and procedural specificities (Mökander et al., 2021; Morley et al., 2021a; Georgieva et al., 2022).

This means that a substantial amount of work remains to be done at the level of individual organisations. They need to identify and tailor suitable sets of tools, structures, and procedures to implement digital ethics in ways that fulfil both high-level principles and specific organisational requirements. The case has thus been made for more support of "bottom-up" AI ethics in the private sector (Mittelstadt, 2019) and the creation of more "worked examples" of how tools have been used to satisfy principles (Morley et al., 2020).

Surprisingly, the role of ethics panels in operationalising digital ethics has received little attention so far, despite their potential to serve as a central hub for the multifaceted efforts required. In other fields, such as business ethics (Schultz & Seele, 2023) and medical ethics (Véliz, 2019), multidisciplinary ethics committees have played crucial roles in developing, implementing and upholding ethical principles, especially when such panels include expertise and experience in ethical reasoning and deliberation (Blackman, 2021). Ethics panels may take an active role in formulating and adapting guidelines and policies suited to their specific organisation, offer practical guidance for their application, and educate both internal and external audiences (Véliz, 2019). They can also integrate different perspectives, both through an interdisciplinary set-up and by engaging with different internal and external stakeholders (Schultz & Seele, 2023).

Yet, while several companies have set up digital or AI ethics boards in recent years, for example, SAP (SAP, 2018), IBM (IBM, 2019) and Orange (Orange, 2021), there is limited information available on how such bodies may work to bridge the operationalisation gap. Only the work of a few digital ethics panels has been submitted to systematic analysis, among them Meta's Oversight Board (Wong & Floridi, 2022) and Microsoft's AETHER Committee (Newman, 2020). Open questions remain as to how ethics panels can make an effective contribution to implementing digital ethics (Sandler et al., 2019; Schuett et al., 2023) rather than, for instance, merely appear as window dressing in "ethics washing" attempts, especially if their recommendations are not binding. There is also surprisingly little information available on how panel

work may be structured to ensure ethical issues are analysed in a fair, reliable and transparent manner, and what can be done to avoid arbitrary decisions or interpretations that are largely influenced by personal intuition (Hirsch et al., 2021).

1.3 The Principle-at-Risk Analysis (PaRA) as an Answer to the Operationalisation Challenge

This paper seeks to address the issue of how ethics panels can contribute to the reliable and transparent implementation of digital ethics in their organisation. It does so by identifying general requirements for effectively connecting ethics panel work to both digital ethics principles and practical problems. We argue that digital ethical panels and companies should be linked through structural measures to avoid the risk that an ethics panel will form an isolated entity within the company. We discuss some best practices on how such links can be facilitated and illustrate them by describing the development of a standardised risk assessment tool, the Principle-at-Risk Analysis (PaRA), to aid the work of the interdisciplinary digital ethics panel at the multinational science and technology company Merck. By sharing how the PaRA tool was devised and providing a worked example of the tool's application in a case concerning medical data sharing, we show how a structured approach to ethics panel work can help to operationalise high-level principles at an organisational level.

We begin by introducing Merck's ongoing efforts to operationalise digital ethics. We then examine operationalisation requirements at the level of a digital ethics panel and present the design of the PaRA tool as a means to guide and harmonise panel work in alignment with relevant ethical principles. Next, we showcase the tool's application in an example case that concerns the comprehensibility of consent forms in a data-sharing context. Finally, we discuss the implications and generalisability of this work and provide an outlook on future steps.

2 Merck and the Digital Ethics Challenge

Originating as a pharmacy in the 17th century in Darmstadt, Germany, Merck has since grown to a large and interdisciplinary multinational organisation, residing in 66 countries, operating across three business lines – healthcare, life sciences and electronics – and employing more than 60,000 people (Merck, 2022). Merck has a strong commitment to responsible entrepreneurship (Oschmann, 2018) and has been proactively seeking ethical guidance for its businesses, most prominently in the biomedical field which, like the digital sphere, tends to progress rapidly beyond the scope of current regulation. Here, the Merck Ethics Advisory Panel for Science and Technology, a group of external academic experts from the fields of biology, medicine, bioethics, philosophy, and law, has since 2011 been guiding the company's healthcare and life science endeavours in ethically sensitive areas of regulatory uncertainty such as genome editing and stem cell research.[1]

---

[1] For an example of the guidance provided by the panel see Sugarman et al., 2018.

Like many other companies, Merck has in recent years undergone significant digital transformation both internally and in its business development. Its digital projects span across drug discovery, supply chain integrity protection and human resources (Merck, 2020), to the services offered by Syntropy, a joint venture between Merck and Palantir that specialises in integrating healthcare data from different sources into a collaborative technology platform for clinical research (Merck, 2020).

Seeking a robust digital ethics strategy, Merck created a digital ethics panel composed of external experts in digital ethics, law, big data, digital health, data governance and patient advocacy (Merck, 2021). The Digital Ethics Advisory Panel (DEAP) meets at least four times a year to discuss ethical issues related to digital projects at the company. Additional ad hoc sessions are scheduled if needed, for example in case of urgent questions. According to its charter, the DEAP is responsible for (1) providing guidance on specific ethical questions related to data and algorithmic systems raised at Merck; (2) evaluating scenarios for ethical risk mitigation proposed by the business units based on panel guidance and offering feedback for their implementation; (3) providing a forum for discussing and evaluating new company policies with ethical impact, as well as Merck's existing ethical principles and policy papers; and (4) proactively advising Merck on relevant developments and emerging areas of discussion on digital ethics.

For its inaugural work, the DEAP developed a Code of Digital Ethics (CoDE) based on a framework of 20 digital ethics principles (Becker et al., 2022). Following analysis of both the academic discourse on digital ethics and a wide selection of existing ethics guidelines, the aim was to establish a versatile pattern-based tool that would lend itself to operationalisation across the full spectrum of the company's digital activities. After adopting the CoDE, Merck charged the DEAP with the dual responsibility of upholding these principles in their own work and ensuring that they are being upheld more broadly across the company's activities.

Although the CoDE was designed with operationalisation in mind, there remains much work to be done to effectively implement it and integrate it into Merck's various workflows. This will require satisfying the requirements outlined in the introduction by committing to design decisions based on the principles set out in the CoDE, establishing repeatable procedures, implementing appropriate oversight, and testing and refining these processes to ensure the reliability of ethical review in different situations and use cases. To roll out the CoDE at all levels of the company, several tools are under consideration. These include basic training for all employees, more advanced courses for teams working with data and algorithms and options for semi-automated ethical risk assessments as part of existing software workflows.

Given the pivotal role of the DEAP in Merck's digital ethics strategy, priority as a pilot project for operationalisation has, however, been given to the development of a procedure that harmonises and guides the DEAP's sessions in alignment with the principles outlined in the CoDE. While ethics boards are as crucial for successful operationalisation (Véliz, 2019; Blackman, 2021; Morley et al., 2021a; Wong & Floridi, 2022; Schultz & Seele, 2023), most methods and tools discussed in Sect. 1 are aimed at individual professionals and teams or describe external auditing methods or project-based checks. They seldom address the specific role of an expert panel and how to best fit its work into the larger effort of applying ethical principles in a com-

plex organisation. We therefore aimed to develop a tool that would be better suited to aiding and standardising the digital ethics panel's work.

3 Putting Guidelines to Work: The Principle-at-Risk Analysis (PaRA)

Based on the preliminary analysis in the previous section, we examined what successful operationalisation of ethics principles might look like at the level of ethics panel work. We identified a need for a framework that allows an ethics panel to consult ethics principles in a standardised yet appropriately contextualised way. The goals of such a framework should be to help the panel conduct structured discussions with consistent links to applicable principles, weigh principles against each other, identify key issues that the company may face if it does not address situations conflicting with the organisation's principles, and propose mitigation measures and other recommendations. We now take a closer look at the requirements that should be met when developing such a tool and illustrate how we implemented these requirements in the development of the Principle-at-Risk Analysis tool as a specific aid for ethics panel work.

3.1 Requirements for Operationalising the Work of Ethics Panels with Ethical Principles

We propose three general requirements for the development of a standardised risk assessment tool. Firstly, it should provide clear cues to structure panel discussions in alignment with an organisation's ethical guidelines, enabling the panel to systematically and reliably consider all ethical principles applicable in a given situation. Secondly, the tool should involve a transparent procedure for selecting and preparing problems, as well as applying ethical principles to specific problems. Thirdly, there should be an infrastructure within the company to provide adequate support to the panel in using the tool and to ensure effective communication between the panel and individual business units.

Concerning the first requirement, it will be helpful if an organisation's digital ethics guidelines are both clearly defined and structured appropriately, so they can easily be considered in turn or in topically related clusters. There should also be a reliable mechanism to quickly identify those principles or guidelines that may be affected by the problem under consideration. Merck's CoDE is already well suited to such operationalisation due to its structured framework, which consists of five core principles and 15 subsidiary principles mapped to these core principles. Such a framework provides a clear understanding of the relationships between principles and their reference to data or algorithmic systems. Each guideline of the CoDE translates one of these principles into statements that aid responsible decision-making in a business environment. However, the guidelines are sufficiently non-prescriptive and technology-agnostic to accommodate different perspectives and the diverse business contexts encountered at Merck (Becker et al., 2022). Such a structure facilitates reliable and systematic consideration of principles by anyone working with the code. Organisations with less structured ethics codes may consider creating additional doc-

umentation to assist structured navigation through their guidelines. Given the limited time of an ethics panel, additional clear mechanisms for identifying those principles that are most likely at risk will also be useful to ensure that the panel can quickly focus on what is important. Merck's Principle-at-Risk Analysis (PaRA) fulfils this by offering a standardised procedure, described in detail in Sect. 3.2, for assessing how principles may be affected or at risk in a given situation.

Regarding the second requirement, a standardised procedure also formalises a transparent process for applying principles to a specific problem. Such a process should meet two objectives: it should guide problem selection and preparation and provide a reliable method for checking a problem against all applicable guidelines. Screening and preparing topics is important to ensure that problems are suitable for the attention of an ethics panel. This involves determining whether a problem constitutes a genuine ethical concern and defining the material scope of the issue (Mökander & Floridi, 2022b; Mökander et al., 2022a). Some issues, for example, may be better addressed as legal or compliance questions, which could be handled by other departments within the company. The PaRA tool meets this first objective by dedicating an initial phase to problem selection and preparation, helping determine if the DEAP should discuss a topic or give priority to one of several suitable issues. To satisfy the second objective, the PaRA provides a formal procedure for checking a problem against all applicable guidelines, determining which principles are affected and at risk of being violated if Merck does not address the stated problem. The details of this procedure are described in Sect. 3.2.

With regard to the third requirement, a dedicated digital ethics support infrastructure will ensure that panel work can proceed effectively and that any recommendations will have the desired impact. A digital ethics service unit can serve as a first port of call for different business units, carry out problem selection and preparation, prepare structured inputs for panel discussions, document the process, and help effectively communicate outcomes. Merck's Digital Ethics (DE) office provides such an infrastructure. The DE office serves as the initial contact point for any enquiries related to digital ethics. Potential issues and problems for consideration by the DEAP are first reviewed by the DE team. They are responsible for implementing the formal parts of the PaRA and for documenting and communicating its results. Having a dedicated DE office has the advantage of having a permanent point of contact for any digital ethics enquiries. The staff in such an office can build up sustainable skills in digital ethics and transfer them to others in the company through their networks. Other organisations may have, or create, different set-ups, such as data ethics or privacy officers within other central company departments or "hub and spokes" networks with staff possessing some digital ethics expertise placed in each business unit (Hirsch et al., 2021). Regardless of the specific set-up, clearly assigning responsibilities for spotting ethical issues, bringing them to a panel and applying a tool such as the PaRA is important to ensure consistent examination of problems in alignment with the applicable guidelines.

The purpose of a tool like the PaRA is thus to guide both those who assess and prepare topics for consideration by a digital ethics panel, as well as the digital ethics panel itself, in ways that ensure close alignment with relevant ethical principles, aid consistency and reduce ambiguity. The tool should provide quality assurance and

standardisation in terms of how principles are assessed. Such a tool cannot make risk identification an automated process or resolve ethical issues on its own. Users of the tool still need to understand problem statements and relevant business contexts and engage in ethical reflection as they consider if and how principles are affected by the case at hand. The PaRA tool was designed with these criteria in mind and aims to spotlight issues related to the digital ethics code, help users systematically consider them, and encourage constructive discussion of digital ethics.

3.2 Design of the Principle-at-Risk Analysis tool

As there was no established method available to link ethics panel work to digital ethics principles according to the requirements identified in the previous section, we took an exploratory approach to design a customised tool to fill this need at Merck. Whilst the specific set-up chosen was tailored to Merck, we believe that a similar approach can work in most other organisations that seek to integrate their digital ethics guidelines more firmly into their ethics panel's work. Based on our analysis, a suitable tool should firstly provide a reliable link between the business units concerned, the support staff of the digital ethics team and the ethics panel. This can be achieved by specifying a clear workflow for the interaction between these groups. Secondly, the tool should be appropriately aligned with the organisation's digital ethics guidelines and enable the panel to discuss issues based on the relevant principles set out in these guidelines. Such an alignment can be created if the tool's questions or checkpoints closely build upon these guidelines and incorporate all their key elements. The advantage of such approach is that the content of the principles of a code is precisely operationalised in the working process.

Given these two criteria, the PaRA tool was designed as an interactive questionnaire based on the content and structure of Merck's CoDE, with several questions developed from the wording of each guideline (see Figs. 2 and 3). It comes with a manual that guides users through the entire process, indicating the required information and outcomes at each step. To ensure the tool's useability and the suitability of its output, the development process was iterative and considered feedback from questionnaire users, panel members and representatives of the involved business to adjust the input sheets, analysis steps and the presentation of the tool's results. Other organisations can follow this example and adapt the workflow outlined below to their own setting.

The PaRA process consists of three phases: (1) preparation, (2) the actual principle-at-risk assessment, and (3) the generation of a report for guiding the digital ethics panel's discussion (see Fig. 1).

During the first phase of a PaRA, the DE team works closely with the business unit that has submitted the enquiry to gather all necessary information and prepare a clear problem statement. The PaRA tool provides a questionnaire to record the initial question as well as relevant details on the technology, business model, likely future developments and customers. The DE team then analyses this information to formulate a problem statement that is both specific and actionable. Statements should be focused on a single topic and revolve around a genuine ethical problem.
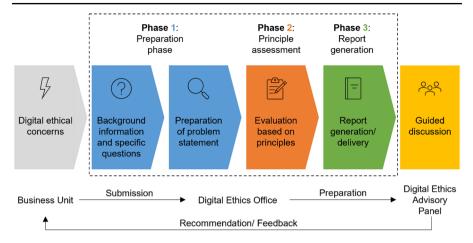
Fig. 1 The Principle-at-Risk Analysis (PaRA) tool

Fig. 2 First stage of the principle assessment – Is a principle affected?

During the second phase of the PaRA process, the DE team scrutinises this problem statement and background information to determine which principles of the CoDE may be at risk in the use case. This principle assessment phase consists of two stages. First, the PaRA asks a series of three basic questions for each of the 15 subsidiary principles outlined in the CoDE to identify any that may be affected. These questions are directly derived from the CoDE's guidelines, each of which corresponds to one digital ethics principle (see Fig. 2).

To avoid every principle appearing affected, thresholds can be introduced. In the context of the PaRA, Merck has decided to consider a principle as affected only if

at least two questions are answered with "yes". This approach recognises that the CoDE's guidelines, on which the questions are based, are formulated at a general level, and as such, there is a high likelihood that at least one question will be answered affirmatively. The drawback of such a threshold is that the analysis may overlook a principle that is severely affected but in only one aspect. To address this issue, users who complete the PaRA questionnaire can manually request further analysis if they believe a principle may be at risk, even if it does not meet the threshold for being considered affected. By offering this option, the PaRA process remains flexible and responsive to the unique circumstances of each use case, ensuring that ethical considerations are always given the attention they deserve.

Once a principle has been marked as affected, the second stage of the principle assessment phase begins. A tool such as the PaRA should provide a thorough yet flexible means to determine if a principle is at risk in the given use case. Careful assessment of potential risks should involve looking at a variety of angles from which a principle's integrity may be compromised. To fulfil this requirement, Merck has developed a series of three or four more detailed questions for each principle to be examined in the second stage of the PaRA's principle assessment. These questions are drawn from the more detailed definitions of the principle behind each guideline, which are also given in the CoDE (see Fig. 3). A principle is considered at risk if at least one out of three, or two out of four, detailed questions are answered with "yes". If this threshold is met, the user is prompted to provide a more detailed explanation for each affirmative answer, along with brief notes on why and how the principle is at risk. Here, too, users can override the threshold and manually highlight a principle "at risk", if necessary.

The third and final phase of the PaRA involves generating an automated report to present the results of the PaRA exercise. The report provides the case summary and problem statement collected in phase one of the PaRA, along with a graphical summary of which of the CoDE's principles are affected and at risk. Additionally, it includes details of all endangered principles and their corresponding core principles, as well as the explanations provided during the PaRA process. For full transparency, an appendix displays the complete set of questions and answers of this particular PaRA. The PaRA report is then used to prepare and guide the subsequent discussion of the DEAP.

We believe that this design of the PaRA serves well to fill the principles-to-practice gap for ethics panel work. An approach such as the one described here satisfies the operationalisation requirements outlined in Sect. 3.1. By delivering a report that clearly shows which principles may be at risk and why, the PaRA helps to structure DEAP discussions in alignment with Merck's CoDE. The tool itself follows a reliably structured procedure for selecting and preparing problems and for applying ethical principles to a specific problem and ensures full and transparent documentation of the process. This encourages the DEAP to take a similarly structured approach in its reasoning and recommendations, thus helping panel members to avoid merely intuitive interpretations and arbitrary decisions. Finally, having clear responsibilities assigned to Merck's DE team for liaising between panel and business units, employing the PaRA tool and documenting subsequent DEAP discussions, ensures the process is carried out reliably with effective communication between all involved parties.

Detailed definition of Privacy

In the context of digital solutions, privacy refers to information about, for example, an individual's family, friends, health and financial circumstances.

Respect for privacy means that each user is able to determine what information is shared with others. This concerns information originally provided by the user as well as information generated about the user in the course of his/her interaction with digital solutions. Otherwise, there is a risk, for example, that users may feel that their privacy is being invaded.

Organisations committed to the principle of privacy respect and guarantee privacy at every stage of the data chain: whether at the beginning when determining the type of processing or at the time of processing within the data chain. This is where measures such as anonymization of personal data, the use of synthetic data or software dashboards to control privacy settings by users come into play.

Since the promotion of the sub-principle of privacy recognises the self-determination of the user, it contributes to preserving the core value of autonomy.

Detail Questions for Privacy

Does the problem statement affect…	yes	no	Explanation (only for "yes") The problem statement affects…	ⓘ
…personalized information about, for example, an individual's family, friends, health or financial circumstances?				
…information originally provided by the user?				
…information generated about the user during his interaction with digital solutions?				
…the need for the protection of personalized user data at any stage of the data chain?				
Principle at risk				

Fig. 3 Second stage of the principle assessment – Is a principle at risk?

4 Example Case: Comprehensibility of Consent Forms for Data Shared by Syntropy

After having outlined the requirements and process for developing a dedicated tool to operationalise digital ethics principles for advisory panel work and having illustrated how Merck has implemented this with the PaRA tool, we would now like to discuss the application of such a tool in more detail. Syntropy was established in 2018 to combine Merck's expertise in healthcare and life science with the software company Palantir's data and analytics platform, Foundry. The purpose of Syntropy is to provide secure sharing and effective analysis of clinical research data. The platform facilitates the structuring and analysis of data from disparate sources, enabling experts from different institutions to collaborate while safeguarding data ownership (Merck, 2018).

Due to its focus on work with sensitive patient data, Syntropy has been an important driver in inspiring the creation of the DEAP and the CoDE. It has also been one of the major sources of enquiries to the DEAP. One of the first problem statements

to be subjected to a PaRA was submitted by Syntropy, addressing ethical concerns about patient consent in a data-sharing context. This example case illustrates how Merck's CoDE was applied to this enquiry with the help of the PaRA tool and how this helped to identify key ethical issues in a concise yet comprehensive manner. It also shows how the PaRA laid the groundwork for a structured and constructive discussion by the DEAP that resulted in tangible recommendations.

4.1 Challenges to Informed Consent in the Digital Era

Informed Consent is a cornerstone of ethical research involving humans (Eyal, 2019). It has found broad uptake in regulations and widely accepted ethical principles in medicine (World Medical Association, 2013; Council for International Organizations for Medical Science, 2002; Office of Human Subjects Research, 2013). The increasing amount of health data being generated and analysed promises significant advances in biomedical research and applications if such data are shared, but this raises questions about the appropriate balance between individual privacy and the public good. (Ballantyne & Schaefer, 2018; Datenethikkommission, 2019).

Recombining data from various sources in large pools and analysing them with the help of increasingly complex algorithms creates novel ethical challenges. For example, it may be difficult to predict at the point of data collection where, how or by whom data will be used. Moreover, the analysis may reveal unexpected findings about patients or their relatives, and data may be susceptible to de-anonymisation, which can lead to individuals being re-identified and exposing sensitive health information to potential abuse. Because of these uncertainties, there is growing concern over third-party-access to health data and widespread aversion to sharing data with health insurance and private pharmaceutical companies (Joly et al., 2015; Patil et al., 2016; Tosoni et al., 2021).

Traditional models of informed consent may be ill-suited to big data projects as they can simultaneously fall short on two fronts. They often fail to both facilitate meaningful data sharing and to ensure adequate data protection and patient autonomy regarding future data use (McKeown et al., 2021; Deutscher Ethikrat, 2017). In addition, legally permissible business practices based on current standards for consent procedures may still be found unethical (Schnebele et al., 2020). For example, consent forms may be hard to comprehend (Manta et al., 2021; Dickert, 2020), lack clarity about how data will be used or shared (Spence et al., 2018) and offer limited options for patients or their surrogate decision makers to express their preferences in more modular or dynamic ways, such as consenting to some types of data use but not others (Spencer et al., 2016).

4.2 Consent Procedures Applied to Data Used by Syntropy

The background information collected during the first phase of the PaRA shows that the Syntropy business model is based on the utilisation of highly sensitive patient data provided by health system institutions. Syntropy enables data curation and aggregation for health system institutions to power their research and analytics. Health system institutions may also offer their data to collaborate with other research institu-

tions or industry partners that require clinical data. These customers pay a license fee for using Syntropy. If a customer engages in a collaboration where payment is made to them, Syntropy may collect a commission. Importantly, Syntropy does not take ownership or control of the data and will only work with or do analysis of the data at the request of the data source provider.

Patients provide consent via the institution or organisation where the data is collected (e.g., health system institutions or research organisations), including general data consent and specific study consent. In the USA, where Syntropy's current data sources are based, general data consent is obtained when patients enter a medical facility for treatment and can be quite broad, to allow the sharing and reusing of deidentified data in a business context, e.g., for billing purposes. There may be little provision for patients to find out afterwards how exactly their data has been used. Consent procedures for specific clinical trials are more stringent and audited by an institutional review board, but even here variability still exists.

Syntropy does not oversee or audit research organisations' patient consent mechanisms. As a result, it is not clear if and in how far Syntropy can proactively contribute to making its business (and patient data use in this business) comprehensible to patients. It is also unclear to what degree Syntropy is accountable for the comprehensibility of consent forms at research organisations and what measures Syntropy should consider based on this level of accountability.

4.3 A Principle-at-Risk Analysis of Syntropy's Approach to Consent Comprehensibility

Given the context presented above, the problem statement submitted into the second phase of the PaRA was: "Should Syntropy care about the comprehensibility of consent?" Putting this problem statement and the background information through the PaRA identified the following principles that are at risk and require further consideration and ethical analysis (see Fig. 4).

Firstly, the core principle autonomy is affected via its subsidiary principles privacy and literacy. Respect for privacy requires that patients can determine what information is shared with others, especially when this involves sensitive medical data, and they can potentially be identified. Literacy is important to ensure patients can make informed and autonomous decisions during the consent process. Patients must be

Justice	Autonomy	Beneficence	Non-Maleficence	Transparency
Impartiality	Explainability	Sustainability	Reliability	Traceability
Equality	Privacy	Security	Controllability	Interactivity
Proportionality	Literacy	Responsibility	Accountability	Comprehensibility

Fig. 4 Overview of the principles at risk identified for the Syntropy use case on patient consent. Affected principles that are also at risk are marked red; those that are affected but not at risk are marked green

able to understand how their data may be used, which in turn requires organisations to take appropriate measures to inform and educate patients about their consent procedures.

Secondly, the core principle non-maleficence is affected via its subsidiary principle accountability. Syntropy's current dependency on business partners to handle the consent process may require the alignment of responsibilities to ensure that accountability is properly distributed and fulfilled.

Finally, the core principle transparency is affected via its subsidiary principle comprehensibility. Valid consent requires patients to understand the processes and implications of data sharing and analysis, which may require communication efforts to address the specific challenges of big data analytics and cater to diverse backgrounds of patients.

The PaRA's results unequivocally support the importance of patient consent and comprehensibility in Syntropy's business model. They provide an affirmative answer to the original problem statement and confirm that these issues require the company's attention. By highlighting which principles are at risk and why, the PaRA provided the necessary context for the DEAP to prioritise ethical principles and explore possible mitigation strategies. As a result, the DEAP's ensuing deliberations focused on formulating actionable recommendations to address these issues.

Ensuring trust is essential for the success of large-scale medical data sharing such as that at the core of Syntropy's business model. Thus, the DEAP discussed the importance of transparency to avoid patient exploitation and reputational damage and to obtain patient buy-in. Transparency is also closely linked, and in many ways a prerequisite, to autonomy (Becker et al., 2022), which is why information is crucial for autonomous patient decisions. Challenges to truly informed consent hinge on transparency as data moves in increasingly complex socio-digital ecosystems that severely reduce the scope for informed decisions at the point of data collection (Deutscher Ethikrat, 2017). Finally, the DEAP identified accountability (as a component of non-maleficence) as a key consideration for potential mitigation measures, including the question of what the specific accountabilities of Syntropy should be.

4.4 Recommendations Developed on the Basis of the Principle-at-Risk Analysis

Syntropy is not currently positioned to interfere with consent procedures carried out by partner organisations that collect patient data. The DEAP therefore explored other ways of ensuring and improving adequate consent procedures and achieving proactive communication about Syntropy's offerings and the way it handles patient data. Following the discussion, the DEAP gave the following recommendations:

Firstly, Syntropy should consider implementing mechanisms for regularly checking the reputational status of its business partners, including both data providers and data consumers, especially before starting a collaboration. This should focus on minimum consent standards and, in addition, data security standards that can be checked before entering a partnership. Secondly, to ensure transparency of consent procedures and promote patient literacy, Syntropy should take proactive steps to engage with the public. This could involve collaborating with patient representatives and patient

advocacy groups such as Friends of Cancer Research or initiating dialogues with stakeholders who may have reservations about digital health.

Although these recommendations are non-binding, Syntropy has embraced them and is currently implementing various activities. To enhance patient engagement, a patient representative is now acting as a guest advisor in all DEAP sessions related to medical contexts. In addition, Syntropy has completed initial work on vetting business partners, by looking into existing practices across the organisation and tools used in the process. Based on this, Syntropy is currently drafting specific vetting procedures and will test and refine them using existing and potential partners as examples. Once finalised, Syntropy will implement these new vetting procedures across the organisation.

Syntropy is actively collaborating with its existing customers and partners to build better consent standards. Their joint efforts are currently focused on developing a partnership-based consent model, in which the patient is regarded as an "educated ally". In recent follow-up work, the DEAP has encouraged the team tasked with this project to further explore and implement the latest guidance and insights from other areas where consent has evolved from being a mere risk management approach to cascading, dynamic or meta-consent models (Loe et al., 2015; Ploug & Holm, 2016; Boers & Bredenoord, 2018; Teare et al., 2021), such as biobanking or non-profit research initiatives (e.g., Count Me In[2], All of Us[3], MIDATA[4]). Building on this foundation, Syntropy will develop initial recommendations and test them with existing and prospective partners, paving the way for the implementation of the new process.

5 Discussion

The Principles-at-Risk Analysis (PaRA) tool was designed to operationalise Merck's CoDE specifically for the work of Merck's DEAP. We believe it offers a promising approach to standardising the application of digital ethics principles for the work of an ethics panel, a context that has received little consideration in the literature until now.

While experts on an ethics panel may be proficient in ethical discourse, ensuring that they apply ethical principles consistently and transparently in ways that are comprehensible to other stakeholders is crucial to supporting and driving operationalisation at other levels of an organisation (Morley et al., 2021a). This task should not be underestimated, even for Merck's DEAP, which is committed to safeguarding the company's digital ethics strategy and was involved in the creation and refinement of the CoDE. Applying the resulting guidelines to specific use cases remains a challenge even in such a set-up. Both the size and thematic breadth of the company and the interdisciplinary nature of the DEAP necessitate care to avoid ambiguities and misunderstandings caused by variations in terminology and traditions across different business sectors or academic disciplines (Schiff et al., 2021). Similar constraints

---

[2] https://joincountmein.org/about-us. Retrieved September 17, 2022.

[3] https://allofus.nih.gov/about. Retrieved September 17, 2022.

[4] https://www.midata.coop/en/home/. Retrieved September 17, 2022.

likely apply in other companies and organisations that work with digital ethics panels. Moreover, even among digital ethics professionals, there is often a tendency to rely on intuition rather than systematically engage with principles (Hirsch et al., 2021). Because a digital ethics panel's discussion time is limited, it is also important to be careful in selecting topics for its attention and to clearly define the meanings and objectives of the issues to be discussed. This is critical to ensure meetings can be effective and address the most pressing concerns.

The PaRA also helps the DEAP to formulate advice that is convincing to recipients. Recommendations based on clearly identified links between use case aspects and the components of the CoDE will be more plausible and easier for business leaders to understand. This increases the likelihood that the DEAP's advice will be taken seriously while reinforcing a shared commitment to the CoDE as a foundation for company decisions. Such effective recommendations and assertiveness are especially desirable when digital ethics panel recommendations are non-binding (Wong & Floridi, 2022). Having a process like the PaRA in place also helps to alleviate concerns around ethics-washing. A tool that structures and documents the application of ethical principles by an ethics panel makes it easier to keep track of the impacts of panel work and follow up on a panel's recommendations. This may not only improve the trust of external stakeholders that an organisation is serious about its ethics panel's work but also assure the advisors who sit on the panel that their work is of value and help to entrench a corporate culture of addressing ethical questions on all levels.

The PaRA tool offers two main benefits. Firstly, it facilitates screening and preparation of topics for the DEAP by providing structured questionnaires to collect background information about each use case and to systematically check if and how each of the CoDE's principles and corresponding guidelines are affected by the problem statement. By following this standardised process, ethical issues relevant to the case can be identified consistently and reliably. These are important steps towards making the digital ethics code actionable. Lack of actionability has been identified as a key shortcoming of many earlier operationalisation attempts (Morley et al., 2020).

Secondly, the PaRA's output provides clear guidance and structure for the DEAP's discussions in alignment with the CoDE. By explicitly linking the issues at hand to the principle definitions in the CoDE, the PaRA report enables constructive discussions that can quickly identify the parties responsible for specific issues and plausible actions to minimise and mitigate the risks. This is particularly important for successful operationalisation and avoiding diffusion of responsibility, which has been highlighted as a key concern (Floridi, 2019; Morley et al., 2021a, b; Mökander et al., 2022a).

The example case presented here shows that the PaRA tool's structured approach has proven useful in guiding the DEAP's discussions of ethical issues related to patient consent collected by clinical centres working with Syntropy. By clearly identifying which of the CoDE's principles were at risk in the given use case, the PaRA report helped the DEAP to quickly hone in on the issues that required the company's attention and to identify potential risks to patient autonomy and potentially unclear levels of accountability among Merck and its partner organisations.

Equally, the focus on principles at risk helped to identify potential solutions: improve the comprehensibility of consent procedures and increase patient literacy.

By displaying detailed explanations of each risk along with the affected guidelines of the CoDE, the PaRA report aided in the development of specific recommendations for improvements based on the commitments made in the CoDE and the constraints of the case, such as Syntropy's limited ability to interfere with the consent procedures of its partner organisations.

Although the PaRA tool was developed specifically to match the structure of Merck's CoDE and the workflow and scope of the DEAP, we believe that there are important general lessons to be learnt and best practices to be devised from the example provided by Merck. The key elements of the PaRA and the general requirements for the development of such a tool that we have outlined here can be adapted to other configurations. Other companies and organisations can learn from our approach and tailor it to match their own specific guidelines, departmental set-up, and digital ethics support infrastructure to support the work of their ethics panel. For example, while the exact content and structure of digital ethics guidelines may differ between organisations, the rationale behind the PaRA can be applied to any set of guidelines. By developing a structured translation of guidelines into questions that can be applied in a transparent and reproducible manner to individual use cases, organisations can aid the work of their ethics panels and the implementation of their recommendations, regardless of the specific set-up of their guidelines.

Our work highlights the importance of having detailed and comprehensive principles and guidelines that are formulated with operationalisation in mind. This was a priority during the development of Merck's CoDE (Becker et al., 2022) and it was crucial in devising a structured approach like the PaRA. Some digital ethics guidelines may not be as readily suited for the development of a tool like the PaRA because they are too "lofty" (Mittelstadt, 2019) and lack sufficient content and guidance to facilitate effective operationalisation. For instance, they may be heterogeneous in their level of reference and mix focus on direct action and organisational elements or be unclear about their underlying principles. Ideally, operationalisation should be considered from the very beginning of the development of a digital ethics code. But, in cases where this has not been possible, organisations can enhance their guidelines with additional details to aid the development of structured tools like the PaRA.

Finally, we highly recommend involving a digital ethics service infrastructure in both the design phase of such tools and their later implementation. Its members can be the first contact point for different stakeholders, operate key stages of tools like the PaRA and contribute to effectively communicating results. A dedicated Digital Ethics (DE) office has worked well at Merck, but other set-ups may work as well, for example digital ethics officers in different business units (Hirsch et al., 2021). The key requirement here is to clearly assign responsibility for spotting ethical issues, bringing them to the attention of the digital ethics panels and other stakeholders, and working with tools such as the PaRA.

6 Conclusion & Outlook

Our work offers insights into how the operationalisation of digital ethics is being implemented on the ground. We discuss requirements for linking ethical principles to the work of an ethics panel and show how Merck has implemented and standardised such a process with its PaRA tool. By sharing the development of this tool along with a practical example of its application, we respond to the demand for more bottom-up ethics in the private sector, more worked examples and more directive approaches (Mittelstadt, 2019; Morley et al., 2020, 2021a). Our focus on the role of an ethics panel, both as a recipient and as a driver of digital ethics operationalisation, addresses a critical area that has until now received little attention but is important for closing the operationalisation gap (Véliz, 2019; Blackman, 2021; Morley et al., 2021a; Wong & Floridi, 2022; Schultz & Seele, 2023). We have closely consulted the current literature on operationalisation throughout this project and hope that the process shared here can serve as a starting point for others looking to apply academic insights to their own digital ethics operationalisation efforts. Moreover, by presenting this work we acknowledge the need to pursue the translation of digital ethics into practice as a dialectical process (Mökander & Floridi, 2021). We explicitly invite constructive feedback and encourage further discussion and research on how to best close the operationalisation gap, both internally and within the wider community.

By actively participating in academic discourse, we also hope to gather valuable insights into the challenges we face as we continue to roll out the CoDE at Merck. Tools like the PaRA can only be one component of the multifaceted approach needed for digital ethics operationalisation (Mökander et al., 2021; Morley et al., 2021a). At Merck, we are currently working on three main areas to further develop our digital ethics operationalisation efforts. These include (1) training employees to integrate the CoDE in their workflows and to recognise potential issues; (2) developing procedures for dealing with problem statements that are not brought to the attention of the DEAP but still require digital ethics advice; and (3) integrating ethics risk assessments into highly automated processes as part of existing software workflows. We will share our insights in these areas to continue advocating for effective and transparent implementation of digital ethics and to create further best practice recommendations that may help other organisations to move their own operationalisation efforts forward.

References

AI Ethics Impact Group (2020, April 1). *From principles to practice: An interdisciplinary framework to operationalise AI ethics*. AI Ethics Impact Group. Retrieved September 12, 2022, from https://www.ai-ethics-impact.org/en

Ballantyne, A., & Schaefer, G. O. (2018). Consent and the ethical duty to participate in health data research. *Journal of Medical Ethics*, *44*, 392–396. https://doi.org/10.1136/medethics-2017-104550

Becker, S. J., Nemat, A. T., Lucas, S., Heinitz, R. M., Klevesath, M., & Charton, J. (2022). A code of Digital Ethics: Laying the foundation for digital ethics in a science and technology company. *AI & Society*. https://doi.org/10.1007/s00146-021-01376-w

Blackman, R. (2020, October 15). A practical guide to building ethical AI. *Harvard Business Review* Retrieved November 29, 2020, from https://hbr.org/2020/10/a-practical-guide-to-building-ethical-ai

Blackman, R. (2021, April 1). If Your Company Uses AI, It Needs an Institutional Review Board. *Harvard Business Review* Retrieved March 25, 2023, from https://hbr.org/2021/04/if-your-company-uses-ai-it-needs-an-institutional-review-board

Boers, S. N., & Bredenoord, A. L. (2018). Consent for governance in the ethical use of organoids. *Nature Cell Biology*, *20*, 642–645. https://doi.org/10.1038/s41556-018-0112-5

Borrego-Díaz, J., & Galán-Páez, J. (2022). Explainable artificial intelligence in data science. *Minds and Machines*, *32*, 485–531. https://doi.org/10.1007/s11023-022-09603-z

Buijsman, S. (2022). Defining explanation and explanatory depth in XAI. *Minds and Machines*, *32*, 563–584. https://doi.org/10.1007/s11023-022-09607-9

Chmielinski, K. S., Newman, S., Taylor, M., Joseph, J., Thomas, K., Yurkofsky, J., & Qiu, Y. C. (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, M. S. A., & Singh, J. (2021). Reviewable automated decision-making: a framework for accountable algorithmic systems. *FaccT'21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* https://doi.org/10.1145/3442188.3445921

Council for International Organizations for Medical Science (2002). *International ethical guidelines for biomedical research involving human subjects*. Council for International Organizations of Medical Science. Re ieved April 24, 2022, from https://cioms.ch/wp-content/uploads/2016/08/International_Ethical_Guidelines_for_Biomedical_Research_Involving_Human_Subjects.pdf

Danish Design Center (2021, December 7). *The Digital Ethics Compass*. Danish Design Center. Retrieved September 12, 2022, from https://ddc.dk/tools/toolkit-the-digital-ethics-compass

Datenethikkommission (2019, October 23). *Gutachten der Datenethikkommission*. Bundesministerium des Innern und für Heimat. Retrieved March 2021, from https://www.bmi.bund.de/SharedDocs/downloads/DE/publikationen/themen/it-digitalpolitik/gutachten-datenethikkommission.pdf;jsessionid=D4EB0659787AC79D9F83548ABF15E36E.1_cid364?__blob=publicationFile&v=7.

Deon (2018). *An ethics checklist for data scientists*. Deon. Retrieved September 12, 2022, from https://deon.drivendata.org/

Deutscher Ethikrat (2017, November 30). *Big Data und Gesundheit – Datensouveränität als informationelle Freiheitsgestaltung* Deutscher Ethikrat. Retrieved March 9, 2021, from https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-big-data-und-gesundheit.pdf

Dickert, N. W., Bernard, A. M., Brabson, J. M., Hunter, R. J., McLemore, R., Mitchell, A. R., Palmer, S., Reed, B., Riedford, M., Simpson, R. T., Speight, C. D., Steadman, T., & Pentz, R. D. (2020). Partnering with patients to bridge gaps in consent for acute care research. *The American Journal of Bioethics*, *20*(5), 7–17. https://doi.org/10.1080/15265161.2020.1745931

Epstein, Z., Payne, B. H., Shen, J. H., Hong, C. J., Felbo, B., Dubey, A., Groh, M., Obradovich, N., Cebrian, M., & Rahwan, I. (2018). TuringBox: An experimental platform for the evaluation of AI systems. *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence* https://doi.org/10.24963/ijcai.2018/851

European Commission (2021, April 21). *Proposal for regulation of the European parliament and of the council – Laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain Union legislative acts* European Commission. Retrieved October 27, 2022, from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206

Eyal, N. (2019, January 16). *Informed Consent*. The Stanford Encyclopedia of Philosophy. Retrieved November 18, 2022, from https://plato.stanford.edu/entries/informed-consent/

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. (2014). *The fourth revolution: How the infosphere is reshaping human reality*. Oxford University Press.

Floridi, L. (2019). Translating principles into practices of digital ethics: Five risks of being unethical. *Philosophy & Technology*, *32*, 185–193. https://doi.org/10.1007/s13347-019-00354-x

Floridi, L., Cowls, J., Beltrametti, 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 and Machines*, *28*, 689–707. https://doi.org/10.1007/s11023-018-9482-5

Floridi, L., Holweg, M., Taddeo, M., Silva, J. A., 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. Retrieved September 11, 2022, from https://forhumanity.center/independent-audit-of-ai-systems

Georgieva, I., Lazo, C., Timan, T., & van Veenstra, A. F. (2022). From AI ethics principles to data science practice: A reflection and a gap analysis based on recent frameworks and practical experience. *AI and Ethics*, *2*, 697–711. https://doi.org/10.1007/s43681-021-00127-3

Gordon, J. (2021). AI and law: Ethical, legal, and socio-political implications. *AI & Society*, *36*, 403–404. https://doi.org/10.1007/s00146-021-01194-0

Grellette, M. (2022). Tech ethics through trust auditing. *Science and Engineering Ethics*. https://doi.org/10.1007/s11948-022-00379-0

Groves, J. (2022, February 8). *Algorithmic impact assessment: a case study in healthcare* Ada Lovelace Institute. Retrieved April 24, 2022, from https://www.adalovelaceinstitute.org/report/algorithmic-impact-assessment-case-study-healthcare/

Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines*, *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 and Ethics*, *1*, 41–47. https://doi.org/10.1007/s43681-020-00008-1

High-Level Expert Group on AI (2019, April 8). *Ethics guidelines for trustworthy AI*. European Commission. Retrieved April 24, 2022, from https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Hirsch, D. D., Bartley, T., Chandrasekaran, A., Norris, D., Parthasarathy, S., & Turner, P. (2021). Business Data Ethics: Emerging Trends in the Governance of Advanced Analytics and AI. Ohio State Legal Studies Research Paper No. 628, Available at SSRN: https://ssrn.com/abstract=3828239 or https://doi.org/10.2139/ssrn.3828239

IBM (2019). Corporate Responsibility Report 2019. Retrieved March 25, 2023, from https://www.ibm.com/impact/files/reports-policies/2019/IBM-2019_Corporate_Responsibility_Report.pdf

Institute for the Future (2018, August 7). *New toolkit shows companies how to anticipate, prevent bad actors from using tech in harmful ways* Retrieved September 12, 2022, from https://ethicalos.org/wp-content/uploads/2018/08/IFTF_Ethical-OS-press-release_8.7.18.pdf

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*, 389–399. https://doi.org/10.1038/s42256-019-0088-2

Joly, Y., Dalpé, G., Derek, S., & Birko, S. (2015). Fair shares and sharing fairly: A survey of public views on open science, informed consent and participatory research in biobanking. *PLOS ONE*. https://doi.org/10.1371/journal.pone.0129893

Keller, S. A., Shipp, S. S., Schroeder, A. D., & Korkmaz, G. (2020). Doing data science: A framework and case study. *Harvard Data Science Review* https://doi.org/10.1162/99608f92.2d83f7f5

Kraus, S., Durst, S., Ferreira, J. J., Veiga, P., Kailer, N., & Weinmann, A. (2021). Digital transformation in business and management research: An overview of the current status quo. *International Journal of Information Management*. https://doi.org/10.1016/j.ijinfomgt.2021.102466

Loe, J., Robertson, C. T., & Winkelman, D. A. (2015). Cascading consent for research on biobank specimens. *The American Journal of Bioethics*, *15*(9), 68–70. https://doi.org/10.1080/15265161.2015.1062174

Manta, C. J., Ortiz, J., Moulton, B. W., & Sonnad, S. S. (2021). From the patient perspective, consent forms fall short of providing information to guide decision making. *Journal of Patient Safety*, *17*(3), 149–154. https://doi.org/10.1097/PTS.0000000000000310

McKeown, A., Mourby, M., Harrison, P., Walker, S., Sheehan, M., & Singh, I. (2021). Ethical issues in consent for the reuse of data in health data platforms. *Science and Engineering Ethics*. https://doi.org/10.1007/s11948-021-00282-0

Merck KGaA (2022, March 3). *Annual Report 2021*. Merck. Retrieved September 15, 2022, from https://www.merckgroup.com/en/annualreport/2021/_assets/downloads/entire-merck-ar21.pdf

Merck KGaA (2018, November 19). *Syntropy to unlock the value of scientific data in fight to end cancer*. Merck. Retrieved November 20, 2019, from https://www.merckgroup.com/en/news/syntropy-19-11-2018.html

Merck KGaA (2021, January 8). *Merck announces formation of Merck Digital Ethics Advisory Panel*. Retrieved February 19, 2021, from https://www.merckgroup.com/press-releases/2021/jan/en/Digital-Ethics-Advisory-Panel-EN.pdf

Merck KGaA (2020, April 14). *Corporate Responsibility Report 2019*. Merck. Retrieved December 11, 2020, from https://www.merckgroup.com/en/cr-report/2019/servicepages/downloads/files/entire_merck_crr19.pdf

Mittelstadt, B. D. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, *1*, 501–507. https://doi.org/10.1038/s42256-019-0114-4

Mittelstadt, B. D., & Floridi, L. (2016). The ethics of Big Data: Current and foreseeable issues in biomedical contexts. *Science and Engineering 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 and Machines*, *31*, 323–327. https://doi.org/10.1007/s11023-021-09557-8

Mökander, J., & Floridi, L. (2022a). From algorithmic accountability to digital governance. *Nature Machine Intelligence*, *4*, 508–509. https://doi.org/10.1038/s42256-022-00504-5

Mökander, J., & Floridi, L. (2022b). Operationalising AI governance through ethics-based auditing: An industry case study. *AI and 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. *Science and Engineering Ethics*. https://doi.org/10.1007/s11948-021-00319-4

Mökander, J., Axente, M., Casolari, F., & Floridi, L. (2022a). Conformity assessments and post-market monitoring: A guide to the role of auditing in the proposed European AI regulation. *Minds and Machine*, *32*, 241–268. https://doi.org/10.1007/s11023-021-09577-4

Mökander, J., Sheth, M., Gersbro-Sundler, M., Blomgren, P., & Flordi, L. (2022b). Challenges and best practices in corporate AI governance: Lessons from the biopharmaceutical industry. *Frontiers in Computer Science*. https://doi.org/10.3389/fcomp.2022.1068361

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. *Science and Engineering Ethics*, *26*, 2141–2168. https://doi.org/10.1007/s11948-019-00165-5

Morley, J., Elhalal, A., Garcia, F., Kinsey, L., Mökander, J., & Floridi, L. (2021a). Ethics as a service: A pragmatic operationalisation of AI Ethics. *Minds and Machines*, *31*, 239–256. https://doi.org/10.1007/s11023-021-09563-w

Morley, J., Kinsey, L., Elhalal, A., Garcia, F., Ziosi, M., & Floridi, L. (2021b). Operationalising AI ethics: Barriers, enablers and next steps. *AI & Society*. https://doi.org/10.1007/s00146-021-01308-8

Newman, J. (2020). Decision points in AI governance. Center for Long-Term Cybersecurity, UC Berkeley. Retrieved September 12, 2022, from https://cltc.berkeley.edu/wp-content/uploads/2020/05/Decision_Points_AI_Governance.pdf

Office of Human Subjects Research (2013). Code of Federal Regulations, Title 45, Part 46. Protection of human subjects. Retrieved September 12, 2022, from https://www.hhs.gov/ohrp/sites/default/files/ohrp/humansubjects/regbook2013.pdf.pdf

Open Data Institute (2021, June 28). The Data Ethics Canvas. Open Data Institute. Retrieved September 21, 2021, from https://theodi.org/article/the-data-ethics-canvas-2021/

Orange (2021). Orange creates a Data and AI Ethics Council. Retrieved March 25, 2023, from https://newsroom.orange.com/orange-creates-a-data-and-ai-ethics-council/

Oschmann, S. (2018). Curious about the future. *Angewandte Chemie Internatiol Edition*, *57*(16), 4108–4109. https://doi.org/10.1002/anie.201711265

Patil, S., Lu, H., Saunders, C. L., Potoglou, D., & Robinson, N. (2016). Public preferences for electronic health data storage, access, and sharing—evidence from a pan-european survey. *Journal of the American Medical Informatics Association*, *23*(6), 1096–1106. https://doi.org/10.1093/jamia/ocw012

Ploug, T., & Holm, S. (2016). Meta consent - A flexible solution to the problem of secondary use of health data. *Bioethics*, *30*(9), 721–732. https://doi.org/10.1111/bioe.12286

Poretschkin, M., Schmitz, A., Akila, M., Adilova, L., Becker, D., Cremers, A. B., Hecker, D., Houben, S., Mock, M., Rosenzweig, J., Sicking, J., Schulz, E., Voss, A., & Wrobe, S. (2021, July 1). *Leitfaden zur Gestaltung vertrauenswürdiger künstlicher Intelligenz*. Fraunhofer IAIS. Retrieved October 10, 2021, from www.iais.fraunhofer.de/ki-pruefkatalog

Priyono, A., Moin, A., & Putri, V. N. A. O. (2020). Identifying digital transformation paths in the business model of SMEs during the COVID-19 pandemic. *Journal of Open Innovation: Technology, Market, and Complexity*. https://doi.org/10.3390/joitmc6040104

Rochel, J., & Evéquoz, F. (2021). Getting into the engine room: A blueprint to investigate the shadowy steps of AI ethics. *AI & Society*, *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. *The International Review of Information Ethics*. https://doi.org/10.29173/irie419

Saleiro, P., Kuester, B., Hinkson, L., London, J., Stevens, A., Anisfeld, A., Rodolfa, K. T., & Ghani, R. (2018). Aequitas: A bias and fairness audit toolkit. https://arxiv.org/abs/1811.05577

Sanderson, C., Douglas, D., Lu, Q., Schleiger, E., Whittle, J., Lacey, J., Nwenham, G., Hajkowicz, S., Robinson, C., & Hansen, D. (2021). AI ethics principles in practice: Perspectives of designers and developers. *Commonwealth Scientific and Industrial Research Organisation*https://arxiv.org/pdf/2112.07467.pdf

Sandler, R., Basl, J., & Tiell, S. (2019). *Building data and AI ethics committees*. Accenture & Northeastern University.

SAP (2018). SAP Becomes First European Tech Company to Create Ethics Advisory Panel for Artificial Intelligence. Retrieved March 26, 2023, from https://news.sap.com/2018/09/sap-first-european-tech-company-ai-ethics-advisory-panel/

Schiff, D., Rakova, B., Ayesh, A., Fanti, A., & Lennon, M. (2021). Explaining the principles to practice gap in AI. *IEEE Technology and Society Magazine*, *40*(2), 81–94. https://doi.org/10.1109/MTS.2021.3056286

Schnebele, C. O., Elger, B. S., & Shaw, D. M. (2020). Google's Project Nightingale highlights the necessity of data science ethics review. *EMBO Molecular Medicine*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7059004/pdf/EMMM-12-e12053.pdf

Schuett, J., Reuel, A., & Carlier, A. (2023). How to design an AI ethics board. https://doi.org/10.48550/arXiv.2304.07249

Schultz, M. D., & Seele, P. (2023). Towards AI ethics' institutionalization: Knowledge bridges from business ethics to advance organizational AI ethics. *AI Ethics*, *3*, 99–111. https://doi.org/10.1007/s43681-022-00150-y

Spence, O., Uba, R. O., Shin, S., & Doshi, P. (2018). Patient consent to publication and data sharing in industry and NIH-funded clinical trials. *Trials*. https://doi.org/10.1186/s13063-018-2651-2

Spencer, K., Sanders, C., Whitley, E. A., Lund, D., Kaye, J., & Dixon, W. G. (2016). Patient perspectives on sharing anonymized health data using a digital system for dynamic consent and research feedback: A qualitative study. *Journal of Medical Internet Research*. https://doi.org/10.2196/jmir.5011

Sugarman, J., Shivakumar, S., Rook, M., Loring, J. F., Rehmann-Sutter, C., Taupitz, J., Reinhard-Rupp, J., & Hildemann, S. (2018). Ethical considerations in the manufacture, sale, and distribution of genome editing technologies. *The American Journal of Bioethics*, *18*(8), 3–6. https://doi.org/10.1080/15265161.2018.1489653

Teare, H. J. A., Prictor, M., & Kaye, J. (2021). Reflections on dynamic consent in biomedical research: The story so far. *European Journal of Human Genetics*, *29*, 649–656. https://doi.org/10.1038/s41431-020-00771-z

Tosoni, S., Voruganti, I., Lajkosz, K., Habal, F., Murphy, P., Wong, R. K. S., Willison, D., Virtanen, C., Heesters, A., & Liu, F. (2021). The use of personal health information outside the circle of care: Consent preferences of patients from an academic health care institution. *Bmc Medical Ethics*. https://doi.org/10.1186/s12910-021-00598-3

UK Statistics Authority (2022, March 31). *Guidelines on using the ethics self-assessment process* UK Statistics Authority. Retrieved July 20, 2022, from https://uksa.statisticsauthority.gov.uk/wp-content/uploads/2021/04/2021_Self-assessment_guidance_V2.3.pdf

UN Global Pulse (2019). *"Risks, Harms and Benefits Assessment". UN Global Pulse*. Retrieved September 11, 2022, from https://www.unglobalpulse.org/policy/risk-assessment/

Véliz, C. (2019). Three things digital ethics can learn from medical ethics. *Nature Electronics*, *2*, 316–318. https://doi.org/10.1038/s41928-019-0294-2

Wachter, S., Mittelstadt, B., & Russell, C. (2021). Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI. *Computer Law & Security Review*. https://doi.org/10.1016/j.clsr.2021.105567

Wong, D., & Floridi, L. (2022). Meta's Oversight Board: A Review and Critical Assessment. *Minds and Machines*, forthcoming, *SSRN*https://ssrn.com/abstract=4255817

World Medical Association (2013). *Declaration of Helsinki. Ethical principles for medical research involving human subjects* World Medical Association. Retrieved September 11, 2022, from https://www.wma.net/wp-content/uploads/2016/11/DoH-Oct2013-JAMA.pdf

Zicari, R. V., Brodersen, J., Brusseau, J., Düdder, B., Eichhorn, T., Ivanov, T., Kararigas, G., Kringen, P., McCullough, M., Möslein, F., Mushtag, N., Roig, G., Stürtz, N., Tolle, K., Tithi, J. J., Van Halem, I., & Westerlund, M. (2021). Z-Inspection®: A process to assess trustworthy AI. *IEEE Transactions on Technology and Society*, *2*(2), 83–97. https://doi.org/10.1109/TTS.2021.3066209

Zorio, S. (2021, April 1). *How a paper by three Oxford academics influenced AWS bias and explainability software* Amazon Science. Retrieved September 12, 2022, from https://www.amazon.science/latest-news/how-a-paper-by-three-oxford-academics-influenced-aws-bias-and-explainability-software