




Ethical Governance of Emerging Digital Technologies in the Public Sector

Insights from Dutch Digital Ethics Commissions

Antonia Sattlegger^(✉) 

Delft University of Technology, 2628 CD Delft, The Netherlands
a.s.sattlegger@tudelft.nl

Abstract. Emerging digital technologies, such as algorithms and machine learning, offer transformative opportunities to public sector organizations but also pose ethical risks and dilemmas. Public sector organizations adopting these technologies need to govern their ethical implications. This research provides insights into the emerging phenomena of digital ethics commissions within Dutch public sector organizations. Composed of external experts on ethics, technology, and governance, these commissions are meant to reflect, advise, and, in some cases, assess the ethical design and use of emerging digital technology and the governance thereof. Through interviews and document analysis, this research explores the motivations, intentions, and perceptions guiding these commissions. Our preliminary findings suggest that, while digital ethics commissions are meant to convey legitimacy, they can also provide external knowledge, open the organizations for reflection from and with society, and contribute to different types of control. In establishing these commissions, government organizations need to balance the formalization of digital ethics governance with the need for a collaborative and reflective ethical practice conducive to (organizational) learning and ethics as contextual practice. This research contributes empirical insights into how public sector organizations address ethical challenges from emerging digital technologies, offering valuable implications for both practitioners and scholars in the field of public administration.

Keywords: Emerging Digital Technologies · Digital Ethics Commissions · AI Ethics · Digital Governance

1 Introduction

Public sector organizations increasingly experiment with and employ emerging digital technologies, particularly those that are data-intensive, such as algorithms and artificial intelligence, to enhance service delivery, improve efficiency, and foster government responsiveness. However, alongside these innovative and transformative opportunities come significant ethical dilemmas and risks. The unethical design and use of emerging digital technologies by public sector organizations not only risks harming individuals

and marginalized groups but potentially undermines trust in government within society at large. The deployment of opaque predictive analytics in welfare service distribution may lead to discriminatory decision-making, limiting human discretion and avenues for redress [1–3]. In particular, core public values, public accountability, and transparency are found to be under threat by the use of algorithms [4].

The ethical design and use of emerging technologies have received more attention, both in practice and research [5–7]. This has resulted in a wealth of design and principle-based ethics codes, guidelines, frameworks, and assessments for emerging technologies, particularly when framed as AI. However, there is a discrepancy between principles and practice. Their impact on the actual design and deployment of AI technologies remains limited [8, 9]. Guidelines primarily focus on theoretical underpinnings rather than offering practical guidance on implementation [10, 11], but remaining “deontological tick-box exercises” [12]. As a result, there’s a growing acknowledgment of the imperative for concrete ethical and socio-legal governance mechanisms to ensure the ethical design and use of emerging technologies by public sector organizations [13]. However, there remains a scarcity of empirical research examining how public sector organizations translate these abstract ethical principles into actionable practices and effectively address ethics in governance practices [14].

We find that public sector organizations in the Netherlands are increasingly establishing digital ethics commissions to facilitate the ethical design and use of emerging digital technologies. These commissions serve as forums composed of external experts in ethics, technology, and governance, tasked with advising on specific cases, themes, and governance issues related to the ethical design, use, and governance of emerging digital technologies. By providing external advice and guidance, these commissions are meant to support public sector organizations in navigating the ethical complexities of emerging digital technologies.

In other domains, commissions dedicated to ethics have been around for longer and, seemingly, stood the test of time. In the healthcare domain, healthcare ethics committees and research ethics committees have been established since 1970. Ethics committees in healthcare first saw a surge in the United States. Highly publicized cases, such as ending life-saving medical treatment of infants born with Down syndrome or clinically dead patients, sparked societal debates around the ethical implications of modern medicine. Subsequent legal cases highlighted the ethical, legal, and professional limitations of courts to address these ethical dilemmas in practice [15]. Emerging technologies have accelerated the possibilities of modern medicine, and these innovative opportunities have given rise to ethical dilemmas. The ambiguous and uncertain nature of decision-making in patient care was exacerbated by (technological) advancements in modern medicine, coupled with societal changes regarding the increasing recognition of individual patient rights and autonomy and the pluralistic clinical and societal context of healthcare institutions [15]. More recent techno-medical innovations, such as genome sequencing and editing, have further exacerbated medical possibilities and respective ethical dilemmas. Medical ethics commissions are to address these uncertainties in research and practice by providing ethical advice inpatient cases and research projects, developing and revising ethical guidelines in applying new technologies, tools, and processes, and providing ethical training to staff. As such, in the medical domain, ethics commissions are part of

broader governance to reduce uncertainties and address ambiguities in the design and delivery of responsible health care [16].

Public sector organizations face similar uncertainties and ambiguities. Emerging digital technologies, such as algorithms, predictive analytics, or large language models, may have a transformative impact on public service delivery by public sector organizations. In this transition period, public sector organizations lack the laws and norms as well as experience in governing emerging technologies and their ethical implications. One may argue that digital ethics commissions could be an instrument to govern the design and use of emerging technologies. Grimmelikhuisen & Meijer [17] emphasize that safeguarding the legitimacy of algorithmic government requires a multiplicity of different institutional arrangements. Such strategies include the strengthening of civic participation and critical thinkers in the design, development, and monitoring of algorithmic systems, which allows “the voices and values of non-technical experts to be given a place in the development of algorithmic systems.” (p. 240). Krijger et al. [18] attribute “the introduction of ethics review boards, codes of ethics, and the engagement of stakeholders in data science processes” as a means to integrate and “promote a culture of ethics and creating awareness for the ethical aspect” throughout the organization, and therefore advanced AI ethics maturity.

Yet, given the heightened public scrutiny of the public sector deployment of emerging digital technologies, PSOs may be tempted to legitimize their actions by a symbolic digital ethics commission. “The setting up of advisory groups that may be powerless or insufficiently critical,” Floridi [19] argues, is indicative of ethics bluewashing. An unethical practice which Floridi [19] defines as “the malpractice of making unsubstantiated or misleading claims about, or implementing superficial measures in favor of, the ethical values and benefits of digital processes, products, services, or other solutions in order to appear more digitally ethical than one is” (p. 187). These attempts, whether intentionally or not, “mask and leave unchanged any behavior that ought to be improved” (p. 187). Ethical principles are decoupled from daily practices and respective outcomes, namely the responsible design of emerging technologies [20]. Therefore, in this research, we address the question of how digital ethics commissions are implemented to contribute to the responsible design and use of emerging digital technologies at public sector organizations.

In this explorative research design, we first conduct a literature review on (digital) ethics commissions, examining their conceptualization and role as translational tools in other domains of applied ethics, particularly corporate AI ethics and biomedical ethics. Second, we detail our methodology, which involves semi-structured interviews, document analysis, and the author’s reflections, to investigate digital ethics commissions in the Netherlands. Fourthly, we present our study’s empirical findings on the perceived needs and motivations for ethics commissions. We provide a definition of digital ethics commissions and distinguish different types of digital ethics commissions according to their impact on the responsible design of emerging digital technologies. Lastly, we discuss our findings, analyzing the contributions and challenges of digital ethics commissions and offering insights for future research and limitations of our study. Our paper aims to contribute to the scholarly discourse on the design, functioning, and implications of digital ethics commissions in public sector organizations, offering empirical insights

for addressing the governance of ethical implications of emerging digital technologies to both academia and practice.

2 Literature Review on Digital Ethics Commissions

Digital ethics commissions could serve as a translational tool between ethical principles and practice and facilitate the ethical design of emerging digital technologies at public sector organizations. In the biomedical domain, similar ethics boards are established in practice. In the (corporate) AI ethics domain, such commissions have only recently appeared. Although the latter has built on the tradition of the former, both domains ascribe different functions and roles to the commissions and, thus, build on different responsibility attributions in applied ethics.

In the biomedical domain, these commissions have been around for a long time. Although they do not address digital emerging technologies specifically, technological innovations, along with societal shifts towards recognizing individual patient rights and autonomy within a pluralistic clinical and societal context of healthcare institutions [15], have given rise to ethical dilemmas, making these commissions seemingly more relevant. Recent techno-medical innovations, such as genome sequencing and editing, have further expanded medical possibilities and heightened associated ethical dilemmas. Two types of ethics commissions can be distinguished [21]: research ethics committees (RECs) and healthcare ethics committees (HECs). RECs have a specified (legal) mandate to review and oversee (medical) research involving human subjects. These committees evaluate the ethical aspects of research protocols, ensuring that the rights, safety, and well-being of the research participants are protected. They assess factors such as informed consent procedures, assessment of potential risks and benefits, data privacy, and compliance with ethical guidelines. These types of research committees have been established at universities and other research organizations to oversee research involving human subjects beyond the medical domain. Healthcare ethics committees (HECs), also known as clinical ethics committees and bioethics committees, are typically established at healthcare institutions involved with patient care. They provide guidance on moral issues relating to patient care, such as moral dilemmas in palliative care. Generally, three functions can be distinguished [15, 16, 22]: (1) ethical case analysis, consultation, and conflict resolution; (2) the development and revision of ethics policies and guidelines, such as patient care protocols; and (3) the education of members, hospital staff, and patients about ethical issues. Despite their maturity, the empirical evidence of their effectiveness is limited and inconclusive [16]. The evaluation of such committees largely relies on subjective measures, such as perceptions of goals attainment, satisfaction and perceived helpfulness of their contributions, particularly, the feeling of relief in moral distress through enhanced understanding and responsibility alleviation. Preliminary success factors are a balance between being embedded in the organization and maintaining a critical independence. Such balance is helped by the multidisciplinary composition of the committees, particularly external expertise in bioethics. Further, these commissions must stimulate collaboration between staff, patients and the committee to stimulate ethical decision-making [16].

In the (corporate) AI ethics literature, digital ethics commissions, also termed AI ethics boards [23], ethics councils [24], ethics advisory board [11] or ethics review

boards [25] haven't made a recent appearance. In reviewing academic and practitioner governance approaches to foster ethical AI, Prem [24] names ethics councils and boards as good practice to provide an infrastructure and communities supporting ethical systems design. Yet, there is little empirical research conducted on these entities [24, 25]. Stahl et al. [25] find that public organizations and private companies apply but a fraction of mitigation strategies discussed in the literature (p. 33), including ethics review boards. Interestingly, they find that ethics review boards feature strongly as theoretical means to increase organizational awareness and reflection. They may enable the engagement with external stakeholder and the reflection on internal processes (p. 34). Similarly, Tiell [26] argues that "committee-based governance" at AI companies can provide valuable insights and feedback to designers, engineers, and executive teams on how the possible impact of AI applications in practice. Schuett et al. [23, 27] propose AI ethics boards as improvements to the corporate risk governance of AI companies, including medium-sized research labs and big tech companies. The authors [25] define an ethics board as "collective body intended to promote an organization's ethical behavior" (p. 1), particularly the reduction of societal risks. Following these authors, the AI ethics board could be attributed several responsibilities to control societal risks: advising the board of directors, overseeing model releases and publications, supporting risk assessments, reviewing the company's risk management practices, interpreting AI ethics principles or serving as contact points for whistleblowers (p. 2). The authors discuss additional design choices, such as the structure of the commission as external or internal body, its composition, decision-making process and mandate (p. 14). Morley et al. [11] argue for an "independent multi-disciplinary ethics advisory board" (p. 251) to be a crucial in providing the core "infraethics" (p. 251). In serving as a translational tool, it addresses above mentioned disconnect between ethical AI principles and the practical design of AI systems. Three tasks can be attributed to the independent advisory board [11]: First, the ethics board is to be instrumental in developing a principle-based ethics code in negotiation with those impacted by the AI system. Second, the board develops a process for the validation, verification, and evaluation of algorithmic designs. Rather than operationalizing the ethics code as "an end-goal that can be objectively achieved, observed, quantified, or compared" (p. 246), the ethics board supports the contextual interpretation of principles and provides further translational tools. Third, the ethics board conducts regular audits of the developed AI systems, the design processes, and the companies ethical conduct at large. Particularly the function of independent oversight requires, according to the authors (2021) a balance between devolved and centralized responsibility attribution in the ethical AI governance through a multi-agent system. The need for algorithmic audits has been emphasized widely. Raji et al. [28] suggest "an internal ethics review board that includes a diversity of voices should review proposed projects and document its view. (...) The purpose of an ethics review board for AI systems includes safeguarding human rights, safety, and well-being of those potentially impacted." (p. 39). Yet, given the lack of accountability and transparency, internal algorithmic audits may not provide the necessary independent assurance and external expertise [29]. Therefore, Morley et al. [11] emphasize the independence of the advisory board providing external interdisciplinary expertise and audit. Moreover, to prevent the

externalization of responsibility for ethical AI, responsibility for the ethical design must be clearly distributed among the independent board and the internal company employees.

Generally, in the (corporate) AI literature digital ethics commissions are implemented as part of the broader formal risk management framework. They are attributed oversight and advisory responsibilities and report to the governing body. This role requires the independence, and thus, distance, from the primary operational processes in the organization. A commission that serves as and contributes to the development of translational tools, such as the development of a principle-based ethics code, remains a top-down, principled approach. This conceptualization differs from the understanding of healthcare ethics commissions in the biomedical domain. As discussed above, the primary responsibility of ethics commissions is to review and advise on specific cases, either patients or research projects. Ethics commissions, often composed by operational staff, such as doctors or researchers, interact with and advise those responsible for daily operations.

3 Methodology

For this research, digital ethics commissions are defined as structured forums advising on the design and use of emerging digital technologies. We included digital ethics commissions if they fulfilled the following criteria: (1) they were an autonomous body with a structured way of working, indicated by, for example, a memorandum or charter, (2) they advised, reviewed or provided oversight on the design and use of emerging digital technologies, (3) by public sector organizations, such as local, regional, and national government and executive agencies. A total of 21 commissions were identified across executive agencies, regional, and local government organizations through a snowball search. Subsequently, we were able to include 15 of these commissions in our data collection. Table 1 provides an overview of the digital ethics commissions included in this research. Multiple commissions were in the process of formation or early implementation stages. The data collection included semi-structured interviews with facilitators of digital ethics commissions, ethics experts and auditors from within the organization, as well as external commission members. The interviews focused on three areas: First, the process and motivation for establishing a commission, its drivers and the needs and objectives the commission was intended to address. Second, the design choices that were made, such as its the composition of the commission, the decision-making process, the mandate and responsibilities. Third, the interview covered questions regarding the embedding of the commission in the existing digital ethics governance and the maturity of said governance. Finally, the interview covered questions on the perceived impact of the commissions, as well as risks and weaknesses. The interviews lasted from 40 to 80 min. The interviews were recorded, transcribed and coded in an open-coding approach. Additional data collection included the gathering of internal and publicly available documents such as official reports, evaluations, and the commission reports on specific cases. These documents provided valuable insights into the activities and decision-making processes of the commissions.

A limitation of our data collection is that we interviewed individuals directly involved with the commissions, rather than those within the organizations responsible for designing the emerging technologies. As a result, our assessment of the impact of these commissions relies on the intentions and perceptions of the interviewees. Furthermore, given

Table 1. Overview Digital Ethics Commissions.

Code	Entity Type	Established	Objective	Focus/ scope of technology	Mandate	Composition	Example cases
C1	Municipality Sounding-board	2022	Reflection on questions posed by the council, the board, the administrative apparatus, explicitly not an advisory body.	Innovative use of data and new technologies.	solicited and unsolicited reflection, non-binding	5 external members	Design and use of dashboards for decision-making, use of cameras and image recognition
C2	Municipality Sounding-board	2022	Case-specific reflection on ethical dimensions and dilemmas	Digital ethics	solicited and unsolicited reflection, non-binding	5 external members	Predictive analytics application
C3	Municipality Sounding-board	2022	Case-specific reflection on ethical dimensions and dilemmas	Innovative technologies	solicited and unsolicited reflection, non-binding	5 external members	Data sharing for the proactive identification of youth with multiple issues. Data aggregation application on neighborhoods to support public investment decisions, Sensors in public spaces, city surveillance cameras, corona tracing app.
C4	Municipality Sounding-board	2021	Case-specific reflection on ethical dimensions and dilemmas	Digital and technological applications involving the collection and processing of data	solicited and unsolicited reflection, non-binding	7 citizen selected on diversity	Automated scan of licenses and parking tickets. Data aggregation in digital twin. Ethical framework for data applications and algorithms.
C5	Municipality Sounding-board	2019	Citizen-participation and reflection on cases and specific questions	Privacy, digital security, and inclusivity	solicited and unsolicited reflection, non-binding	Self-selected group of citizen	3D mapping via satellite images, COVID-19 Crowd Density Monitoring App

(continued)

Table 1. (continued)

C6	Regional Sounding-board	2023	Theme-based reflection on ethical dimensions and dilemmas	Digital technology, algorithms, data applications	solicited and unsolicited reflection, non-binding	7 external members	Advice on the responsible use of ChatGPT
C7	Executive agency Sounding-board	2022	Theme-based reflection on ethical dimensions and dilemmas	New technologies	solicited reflection, non-binding	10 external members	Reflection on topics, such as ethical framework for facial recognition technology, and ethical framework for internet surveillance
C8	Municipality Advisory Committee	2023	Case-specific advice on ethical dimensions and dilemmas, and advice on the development of a value framework	Data and technology	solicited and unsolicited reflection, non-binding	5 external members	Just started
C9	Municipality Advisory Committee	2022	Case-specific advice on ethical dimensions and dilemmas, with particular focus on the governance questions	The use of algorithms, particularly algorithmic governance	solicited and unsolicited reflection, non-binding	9 external members	Facial recognition upon registration (procured software), Register for sensors in public spaces
C10	Municipality Advisory Committee	2018	Case-specific advice on ethical dimensions and dilemmas	Technology and innovation	solicited and unsolicited reflection, non-binding	7 external members	surveillance in high-risk areas within the municipality, app to gain better insight into night-time street harassment, advice on Principles of the Digital Society
C11	National Advisory Committee	2023	Case-specific advice on ethical dimensions and dilemmas	Data analysis, algorithms, risk models, artificial intelligence	solicited and unsolicited reflection, non-binding	11 external members	Just started
C12	Municipality Supervision Committee	2020	Case-specific advice on ethical desirability	Complex and/or politically sensitive personal data processing technology	solicited and unsolicited reflection, formal comply-or-explain, official responsible reports deviation to council and	7 external members	Risk model for welfare support applications, algorithmic bias analysis

(continued)

Table 1. (continued)

C13	Executive agency Supervision	2022	Case-specific ethical judgement	Data applications	solicited, comply-or-explain	Multiple internal members	Provision of data and data protection
C14	Executive agency Supervision	2021	Case-specific ethical judgement	Data research projects using personal data	solicited, binding	Multiple internal members	Biases in data research projects
C15	Executive agency Advisory/Supervision	2022	Case-specific advice on ethical dimensions and dilemmas	Personal data analysis, algorithms, predictive analytics, artificial intelligence with impact on citizens.	solicited and unsolicited reflection, non-binding	7 internal and 2 external members	Fraud prediction algorithm, proactive service based on vulnerability prediction algorithm

that digital ethics commissions are a relatively recent phenomenon, with most established within the last two years, their mid- and long-term impact remains to be assessed. A potential limitation arises from the involvement of two of the authors with different commissions at an executive agency. While these commissions were not formally included in the research, the authors' experiences may have influenced their reflections on ethics commissions. To mitigate this risk, shared reflection and critical analysis were employed throughout the research process. This research was motivated and informed by the authors' practical experience with digital ethics commissions in the Netherlands. The first author is involved as member of a digital ethics work group facilitating an ethics commission at an executive agency (C15). The second author is active member of two ethics commissions (C6 & C9). This involvement and the experience in working with ethics commissions from different perspectives is a valuable insight. This personal connection also introduces a personal bias. We seek to mitigate this risk by shared reflection and critical analysis through comparing our experiences to other commissions.

4 Results

In the following section, we present our empirical results by addressing three key questions: First, we identify the perceived challenges and needs in the adoption of emerging digital technologies which motivated the establishment of these commissions. Second, we analyze how these commissions are meant to address these needs. By deducing shared characteristics of these commissions, we arrive at a definition of digital ethics commissions as employed by Dutch public sector organizations. Third, we explore the different ways digital ethics commissions are intended to contribute to the responsible design and use of emerging digital technologies, culminating in a typology of their various designs.

4.1 What are Digital Ethics Commissions?

To address these needs, public sector organizations have, or are in the process of, establishing digital ethics commissions. We find that digital ethics commissions can be defined as structured forums advising on the design and use of emerging digital technologies. Table 1 provides an overview of the commissions we identified in the Netherlands. While the design of commissions varies, we find that they generally share three characteristics.

Scope. Digital ethics commissions address ethical challenges and dilemmas related to the adoption and governance of emerging digital technologies. Particularly sounding boards employ a broader scope, including technologies such as open data platforms, digital twins, geospatial technologies and smart city technology, such as drones, surveillance cameras, or crowd control. Supervision commissions adopt a narrower scope, typically addressing algorithms and artificial intelligence. The working definition of these technologies is subject to discussion and tends to become more encompassing with the increased maturity of the commission.

Structured Forums. These commissions are composed of a purposefully selected group of members, have a structured way of working and a formal mandate. This organization gives the commissions a certain degree of agency and allows them to engage in a shared learning and development process with the organization, differing from the ad-hoc participation of citizen or consultation of experts. Commissions generally have a specified formal mandate to provide advice on ethical issues, though the strength of this mandate varies, as we will discuss below. This advice is usually provided in a written report which reflects the reflection and assessment of the commission members. Generally, commissions take a concrete case as their starting point, although some are also reflecting on broader questions. In those commissions who work with a case-based approach and prepare specific questions to deliberate with the commission, both the facilitators and the members seem to be more positive about the useability and impact of the advice. As one member argues: “it is often the application of the technology in the context that really provides depth. So what values are really at stake here? So that was really looking at it from a case-by-case perspective, which also makes the conversation more valuable and ultimately the yield as well.” (C3F1). Another commission member agrees “We also looked at how they deal with an algorithm register? How do they ensure they have enough overview and which algorithms are all within the Municipality? The cases are still a kind of entry point to discuss broader policy. So ultimately, the cases are not the core for us. But my experience is that if you only talk about policy in abstract terms, then you actually have no idea how that works out in concrete practices. Practice is always also an entry point for a broader kind of consideration.” (C9M1).

Degree of Independence. The independence of the commission is repeatably emphasized by the interviewees and the documentation of the commissions. Though degrees may vary, this independence lies in the external composition of the commission, its positioning as autonomous body and its mandate to give solicited as well as unsolicited advice. Although the ability to provide unsolicited advice seem to play a rather ideational role, as we could find only two instances of unsolicited advice. The commissions are external by selecting members from outside the organization. Commissions are composed of external experts with a background in (applied) ethics, computer science, public

administration, business analysts, and legal experts. Particularly university professors are perceived as conveying a high degree of independence. Other commissions are composed of citizen with or without specific knowledge in the domain. Both compositions are meant to provide an external and independent view on the organization's inner workings, as "view from the outside" (C4F1) by "people from other worlds. With a different perspective on things." (C9F1). Other commissions consist of internal members or a mix of internal and external members. In these cases, interviewees emphasize the independence of the commission's advice by highlighting the autonomy granted by its mandate and self-governing capabilities.

4.2 Which Needs are Digital Ethics Commissions Meant to Address?

Digital ethics commissions are generally meant to address four interrelated challenges which arise with emerging digital technologies: the perceived complexity, uncertainty, and ambiguity related to the adoption of emerging digital technologies, and socio-political pressures, such as the lack of societal trust in the use of emerging digital technologies, particularly algorithms and artificial intelligence.

Complexity. Interviewees experience various types of complexity, such as technological, epistemic, and organizational complexity. For one, complexities relating to the digital technology and its use. This is exemplified by the discussions and the perceived need for a shared understanding of (technical) concepts, such as algorithms or artificial intelligence. This complexity is related to the "interweaving of technology and ethics and other governance issues. These questions involve intertwining backgrounds, knowledge of technology, but also knowledge of law or governance. Or ethics." (C9M1). In their perception, public organizations are particularly unequipped to deal with such interconnected complexities, which requires cross-silo thinking. This "interweaving" requires the involvement of multiple organizational entities such as the FG, the CIO, the CDO, in some cases an ethics officer. This creates a "complex field of forces surrounding it" (C9M1).

Uncertainty. Interviewees also express uncertainty regarding the emerging and transformative character of these technologies. For one, technology is rapidly evolving, as well as the possibilities and use-cases it affords. For another, interviewees express uncertainty about the possible impact on citizens and society. The transformative impact of these technologies is expected to be great, and so are the associated risks – "you still have a kind of yes unknown territory where the risks are estimated to be high." (C9M1). Due to the novelty and emerging character of these technologies, there is a lack of experience, norms and regulation governing them. One interviewee summarizes "The laws and regulations are lagging behind, so we have to do other things to ensure that it is used in a responsible manner." (C11M2). While this delay is perceived in terms of time, stricter regulation is also not always desirable, as one participant argues: "And you always have that consideration that you don't want to be too restrictive because then you might hinder certain innovations. Sometimes it's just very complicated to see when new technology comes into society and what the negative effects of that actually are." (C10M1).

Ambiguity. Interviewees also express ambiguities in the application of these technologies and the opportunities they afford. This is exemplified by an often-raised dilemma between what is perceived as desirable and what is legally possible in the context of using personal data for bias testing. Interviewees perceive these value conflicts to be particularly pressing in the context of public organizations, as they not only have to realize moral values through their design artifacts, such as fairness, but also public values through the artifact, the process and organization. This ambiguity is often not reflected in ethical design principles, C7M1 argues.

Transition Period. Facing these challenges, many interviewees recognize that public sector organizations find themselves in a transition period. This transition period is characterized by the above-mentioned emerging challenges. Together with a developing governance structure for emerging technologies, and particularly the governance of ethical aspects. When asked to assess their organizations maturity according to the AI ethics maturity framework by (Krijger et al., 2023), interviewees assess the maturity as generally low, particularly on the level of “orientation on frameworks, guidelines/principles, trainings on data science ethics takes place in teams”. Overall, the organizations find themselves in a “search” (C2F1) or “learning process” (C8F1), and they “are just not there, yet” (C3F1) to effectively address the ethical implications of emerging digital technologies.

Lack of Legitimacy and Trust. The use of predictive algorithms, as one particular type of emerging digital technologies, by public sector organizations in the Netherlands is under increased public scrutiny, especially in the social welfare domain. A prominent case involved the Dutch tax authorities using a predictive algorithm to wrongfully accuse childcare benefits recipients, primarily from minority backgrounds, of fraud. This has heightened public sensitivity and generated distrust towards the government’s use of algorithms. All but two interviewees mentioned this case to illustrate the importance of digital ethics (commissions) at one point of the interview. Public sector organizations are perceived to be standing “under a magnifying glass” (C3F1) – doing “anything with algorithms, you are already on the backfoot” (C15M1). The increased political sensitivity and distrust contributes to a need of legitimization of the design of algorithms. Digital ethics commissions have been established largely at the initiative of political-administrative leadership. Respectively, the need for digital ethics commissions is not always shared. As one interviewee remarks, “I wasn’t waiting for it (the digital ethics commission),” states an interviewee, “we were already doing well with an internal working group.” (C8F1). While there may be various means to legitimize the design of algorithms, digital ethics commissions are particularly widely applied and discussed in the landscape of government organizations. Multiple interviewees raise mimetic and normative isomorphism [30] as drivers for the establishment of digital ethics commissions across government organizations. “Every municipality that wants to be taken seriously now needs an ethics commission. Municipality [X] has one, so does municipality [Y], then we need one too. That is just how it goes. It’s become a trend.” (Interviewee).

4.3 What are Different Design Choices of Digital Ethics Commissions that are Made to Address These Challenges?

Government organizations design digital ethics commissions differently to address these needs. The need to address the internal challenges related to the complexity, uncertainty and ambiguity and the external socio-political pressures arising with the adoption of emerging digital can be either addressed along an important continuum. On the one end, the formalization and assurance of digital ethics governance through the digital ethics commissions. On the other end, the facilitation of external ethical reflection and collective engagement in promoting and developing ethical development. We find that commissions develop from their initial intention and throughout their implementation and may be understood as different types. Generally, we find that three types of ethical commissions can be defined along this line:

Sounding board – A sounding board provides ethical reflections on topics and governance put forward by the facilitator or working group.

Advisory committee – An advisory committee provides actionable recommendations on dealing with ethical dilemmas in specific projects and governance aspects. An advice can have a formal or informal comply-or-explain mandate.

Supervision commission – A supervision commission provides a judgement as to whether it is perceived as ethical to proceed with a project.

While there is variance both across the commissions as well as in their development, we find the following characteristics to be indicative for these different types of commissions – and the respective understanding of governing and implementing digital ethics.

Mandate – Three types of mandates can be differentiated: First, a reflection by sounding boards, provides identifications and reflections on ethical dilemmas, or “the uncovering of certain tensions” (C11M2). Reflections are formulated to avoid recommendations in terms of actionable guidance or course for action. As one commission member elaborates, “(...) Our reports are a contribution to the political debate and are not final judgment. It is rather a starting point for a kind of joint reflection. Second, an advice by advisory boards, which provides recommendations on whether or not a technology or certain course of action is ethically desirable, as well as assessments on moral values and dilemmas. To arrive at an advice, commissions often, but not always, use an assessment framework, such as a catalogue of organizational core values. An advice is non-binding and rarely supported by a formal comply-or-explain procedure. Third, a judgement by a supervision commission, which entails a clear directive as to whether to proceed or not with a certain project. There is only one organization whose commissions judgements has a strong comply-or-explain mandate.

Addressee – Closely related to the mandate, we find that digital ethics commissions vary as to the actors they aim to address, engage with and impact. Supervision commissions address case owners, such as the designers and developers with their ethical judgments. While cumulative reports are extended to administrative leadership, case owners are responsible for approaching the commission and implementing the ethical judgment. Sounding boards seek to address a broader audience, particularly political-administrative leadership. One commission member understands the commission as “an

attempt at democratizing ethics. (...) On the one hand, introducing ethical arguments in a flexible way into the democratic debate in the City Council. So, enabling councilors to have as broad an ethical discourse as possible about new technology. (...) And on the other hand, empowering citizens to ask those ethical questions in a much broader way themselves. And I think that's really the strength of such an approach." (C3M2). Advisory committees generally address both case owners and administrative leadership. Interestingly, despite their varying mandates and addressees, digital ethics commissions generally extend their reflections as written reports. Few commissions engage or experiment with different forms which could be more conducive to shared reflection and learning. Two advisory boards emphasize the importance of shared sessions between the commission and internal working groups. Another, supervision commission organizes annual panels as means of interacting with political-administrative leadership.

(Dis-)integration in the organization – Lastly, the integration and disintegration of the ethics commission form the organization is indicative for the different types of commissions. Supervision commissions are most closely integrated and formalized in the organizational processes and structures. Often, they are understood to be part of the organizational risk management or are connected to existing privacy governance. Sounding boards are the most removed from the organization. This is exemplified by how cases are referred to the commissions. Particularly, supervision boards, and, to increasingly advisory boards, have a formal process of referring cases to commissions. In few instances organizations have established a risk scan, which indicates when case owners are to engage with the commission. In other cases, cases are escalated through the internal working group to the commission, such as, if a certain dilemma is perceived to be sufficiently “challenging”, “interesting” or “representative” by the facilitator. Again, advisory boards fall in between and, therefore, interviewees at times express a tension between the independent mandate of the commission and the embedding in the organization.

5 Discussion, Conclusion and Limitations

In this research we explore the question as to how digital ethics commissions are implemented to contribute to the responsible design and use of emerging digital technologies at public sector organizations. We find that digital ethics commissions are addressing both design and use challenges related to the degree of complexity, uncertainty and ambiguity which emerging digital technologies introduce, as well as a need for assurance and oversight related to the perceived lack of legitimacy and public trust in the governments use of emerging digital technologies, particularly algorithms and artificial intelligence. Public sector organizations address these needs differently through varying implementations of digital ethics commissions – sounding boards, advisory committees and supervision commission. Digital ethics commissions may address these needs by providing external knowledge and reflection. They may also serve as assurance and oversight body, which relates to the functions that commissions are attributed in biomedical ethics and (corporate) AI ethics. Digital ethics commissions can also be indicative for a balance public sector organizations are seeking between the formalization digital ethics governance instruments, procedures and practices in bureaucratic organizations, on the

one hand, and facilitating collaborative ethical reflection and organizational learning on the other hand.

While we have analyzed the motivations, intentions, and perceptions behind the establishment of these commissions, we have not evaluated the actual impact on design and use of emerging digital technologies by public sector organizations. This limitation should be addressed through further research into the actual impact of ethics commissions in on actual organizational practices and their materialization in the design of technology. Such research should include those addressed by the commissions, particularly project owner and managers, and administrative as well as political leadership. This is another limitation of this study. Our interviews were limited to facilitators of the commissions. Though these individuals are most knowledgeable in the workings of such commissions, they personal stake in the commissions could introduce a bias. A bias could also be introduced through the authors involvement in the work of ethics commissions, as we have previously discussed.

This paper offers preliminary empirical insights into how public sector organizations address the ethical challenges posed by emerging digital technologies through digital ethics commissions. By analyzing the motivations, intentions, and perceptions guiding the establishment of digital ethics commissions, it illustrates how government organizations seek to balance the need for organizational digital ethics maturity and formalization, on the one hand, and the need for shared reflection and learning in a transition period.

References

1. Margetts, H.: Rethinking AI for good governance. *Daedalus* **151**, 360–371 (2022). <https://doi.org/10.2307/48662048>
2. Peeters, R., Widlak, A.C.: Administrative exclusion in the infrastructure-level bureaucracy: the case of the Dutch daycare benefit scandal. *Public Adm. Rev.* **83**, 863–877 (2023). <https://doi.org/10.1111/PUAR.13615>
3. Peeters, R., Widlak, A.: The digital cage: administrative exclusion through information architecture – the case of the Dutch civil registry’s master data management system. *Gov. Inf. Q.* **35**, 175–183 (2018). <https://doi.org/10.1016/j.giq.2018.02.003>
4. Busuioc, M.: Accountable artificial intelligence: holding algorithms to account. *Public Adm. Rev.* **81**, 825–836 (2021). <https://doi.org/10.1111/puar.13293>
5. Tsamadou, A., Aggarwal, N., Cowls, J., et al.: The ethics of algorithms: key problems and solutions. *AI Soc.* **37**, 215–230 (2022). <https://doi.org/10.1007/s00146-021-01154-8>
6. Buijsman, S., Klenk, M., van den Hoven, J.: Ethics of Artificial Intelligence
7. van de Poel, I.: Embedding values in Artificial Intelligence (AI) systems. *Minds Mach. (Dordr)* **30**, 385–409 (2020). <https://doi.org/10.1007/S11023-020-09537-4>
8. Jobin, A., Ienca, M., Vayena, E.: Artificial Intelligence: the global landscape of ethics guidelines. *Nat. Mach. Intell.* **1**, 389–399 (2019)
9. McNamara, A., Smith, J., Murphy-Hill, E.: Does ACM’s code of ethics change ethical decision making in software development. In: *Proceedings of the 26th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE 2018)*, Lake Buena Vista, FL, USA, 4–9 November 2018. ACM, New York (2018)
10. Morley, J., Floridi, L., Kinsey, L., Elhalal, A.: 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 (2020). <https://doi.org/10.1007/s11948-019-00165-5>

11. Morley, J., Elhalal, A., Garcia, F., et al.: Ethics as a service: a pragmatic operationalisation of AI ethics. *Minds Mach. (Dordr)* **31**, 239–256 (2021). <https://doi.org/10.1007/s11023-021-09563-w>
12. Hagedorff, T.: The ethics of AI ethics: an evaluation of guidelines. *Minds Mach. (Dordr)* **30**, 99–120 (2020). <https://doi.org/10.1007/s11023-020-09517-8>
13. Theodorou, A., Dignum, V.: Towards ethical and socio-legal governance in AI. *Nat. Mach. Intell.* **2**, 10–12 (2020)
14. Sigfrids, A., Nieminen, M., Leikas, J., Pikkuaho, P.: How should public administrations foster the ethical development and use of artificial intelligence? A review of proposals for developing governance of AI. *Front. Hum. Dyn.* **4**, 858108 (2022). <https://doi.org/10.3389/FHUMD.2022.858108/FULL>
15. Aulisio, M.P., Arnold, R.M.: Role of the ethics committee: helping to address value conflicts or uncertainties. *Chest* **134**, 417–424 (2008). <https://doi.org/10.1378/CHEST.08-0136>
16. Crico, C., Sanchini, V., Casali, P.G., Pravettoni, G.: Evaluating the effectiveness of clinical ethics committees: a systematic review. *Med. Health Care Philos.* **24**, 135–151 (2021). <https://doi.org/10.1007/S11019-020-09986-9>
17. Grimmelikhuijsen, S., Meijer, A.: Legitimacy of algorithmic decision-making: six threats and the need for a calibrated institutional response. *Perspect. Public Manag. Gov.* **5**, 232–242 (2022). <https://doi.org/10.1093/ppmgov/gvac008>
18. Krijger, J., Thuis, T., de Ruiter, M., et al.: The AI ethics maturity model: a holistic approach to advancing ethical data science in organizations. *AI Ethics* (2022). <https://doi.org/10.1007/s43681-022-00228-7>
19. Floridi, L.: Translating principles into practices of digital ethics: five risks of being unethical. *Philos. Technol.* **32**, 185–193 (2019). <https://doi.org/10.1007/s13347-019-00354-x>
20. Meyer, J.W., Rowan, B.: Institutionalized Organizations: Formal Structure as Myth and Ceremony (1977)
21. Steinkamp, N., Gordijn, B., Borovecki, A., et al.: Regulation of healthcare ethics committees in Europe. *Med. Health Care Philos.* **10**, 461–475 (2007). <https://doi.org/10.1007/s11019-007-9054-6>
22. Hajjibabae, F., Joolae, S., Cheraghi, M.A., et al.: Hospital/clinical ethics committees' notion: an overview (2016)
23. Schuett, J., Reuel, A.-K., Carlier, A.: How to design an AI ethics board. *AI Ethics* (2024). <https://doi.org/10.1007/s43681-023-00409-y>
24. Prem, E.: From ethical AI frameworks to tools: a review of approaches. *AI Ethics* (2023). <https://doi.org/10.1007/s43681-023-00258-9>
25. Stahl, B.C., Antoniou, J., Ryan, M., et al.: Organisational responses to the ethical issues of artificial intelligence. *AI Soc.* **37**, 23–37 (2022). <https://doi.org/10.1007/s00146-021-01148-6>
26. Tiell, S.: Create an Ethics Committee to Keep Your AI Initiative in Check. *Business Ethics* (2019)
27. Schuett, J.: Three lines of defense against risks from AI. *AI Soc.* (2023). <https://doi.org/10.1007/s00146-023-01811-0>
28. Raji, I.D., Smart, A., White, R.N., et al.: Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing. In: *FAT* 2020 - Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pp. 33–44. Association for Computing Machinery, Inc. (2020)
29. Mökander, J., Floridi, L.: Ethics-based auditing to develop trustworthy AI. *Minds Mach.* **31**, 323–327 (2021). <https://doi.org/10.1007/s11023-021-09557-8>
30. Dimaggio, P.J., Powell, W.W.: The iron cage revisited: institutional isomorphism and collective rationality in organizational fields. In: *The New Economic Sociology*, pp. 111–134 (2022). <https://doi.org/10.1515/9780691229270-005/PDF>