The Underlying Values of Data Ethics Frameworks: A Critical Analysis of Discourses and Power Structures

Abstract: A multitude of ethical guidelines and codes of conduct have been released by private and public organizations during the past years. Those abstract statements serve as a response to incidents of discriminatory algorithms and systems and have been quantitatively investigated for the proclaimed principles. The current study focuses on four frameworks designed for application during the development of new technologies. The purpose is to identify values and value conflicts and consider how these are represented in relation to established discourses, practices, and attitudes in Computer and Information Ethics. This helps to understand to what extent the frameworks contribute to social change. Critical Discourse Analysis according to Fairclough is used to examine language and discourses, and review edition and publication processes. Well-established values like transparency, non-maleficence, justice, accountability, and privacy were detected whereas value conflicts were barely addressed. Interestingly, the values were more often framed by a business, and technology discourse than an ethical discourse. The results suggest a hegemonic struggle between academia and tech industry whereas power asymmetries between developers and stakeholders are reinforced. It is recommended to extend stakeholder participation from the beginning and emphasize value conflicts. This can contribute to advance the field and effectively encourage a public debate about the desired technological progress.

Keywords: critical discourse analysis, data ethics framework, human values, information ethics

1 Introduction

During the last decade, numerous headlines and (news) articles about discriminating algorithm decision-making, harmful applications or data breaches have shown the drawbacks of technological progress. Those incidents have raised awareness for the impact of algorithmic bias and set ethical questions on the public agenda (Fast and Horvitz 2016). Since 2016, a growing number of reports deals with the subjects of artificial intelligence (AI) and ethics (Whittlestone et al. 2019b). Around the same time, many private and public organizations addressed the concern towards harmful technologies by releasing codes of ethics and guidelines. Those declarations are non-enforceable soft law codes and usually voluntary commitments or general recommendations (Haas and Gießler 2020). The increasing number of codes and guidelines manifests the organizations’ interest in shaping ethical guidance according to their preferences (Jobin, Ienca, and Vayena 2019) but also leads to accusations of “ethics washing” (Floridi 2019).

However, the codes of ethics proved to be rather abstract and not helpful in guiding ethical decision making in daily work tasks (Morley et al. 2020; O’Boyle 2002). Lately, there have been a few publications that are explicitly designed for practical application (e.g. Center for Democracy & Technology 2017; DataEthics.eu 2021; Department for Digital, Culture, Media & Sport UK 2018; Schäfer and Franzke 2020; Tarrant and Maddison 2021). In this paper, they will be referred to as Data Ethics Frameworks (DEF) and are meant to address ethical issues during the early stages of development and design of a technological system or service. The DEFs are targeted not only at engineers and developers but to the persons who are involved in a specific project which makes use of data or applies algorithmic systems. The frameworks’ appealing design and their brief questionnaire genre facilitate the ethical deliberation process and are the major differences compared to conventional guidelines and codes.

The purpose of the current study is to identify the values and value conflicts conveyed via the DEFs. Applying the Critical Discourse Analysis (CDA) according to Fairclough (1995) to four frameworks allows to further
investigate how the values and value conflicts are presented and which discursive patterns are used. In this respect, DEFs are considered an expression of the social practices in the Computer and Information Ethics (CIE) discipline and thus are shaped by the corresponding beliefs and processes. Consequently, the relations and power structures between the involved actors and stakeholders should be critically examined. Finally, this holistic approach enables discussing the frameworks’ contribution to social change. In a matter that is developing so rapidly and changing our society so profoundly, it is of particular interest to know what motives prevail and who is taking part in shaping the discourse.

The originality of the study is obtained by the qualitative approach and the chosen methodology since ethical guidelines have been analyzed predominantly quantitatively with content analysis (Hagendorff 2020; Jobin, Lenca, and Vayena 2019; Schiff et al. 2021) or with frame analysis (Greene, Hoffman, and Stark 2019). Considering value conflicts as well as practices and attitudes of the social field achieve an amplified understanding of issues at stake. Power asymmetries were discussed on a broad level in previous studies (Jobin, Lenca, and Vayena 2019; Schiff et al. 2020). Thanks to a methodology that is sensible towards ideological language and hegemonic views from the perspective of oppressed groups, new insights on power structures are added.

This paper first gives a brief overview of the foundations of CIE and currently discussed subjects. The second part introduces the CDA methodology and the proceeding of the four analyzed frameworks. Subsequently, the identified values, value conflicts, and discourses are reported, following a discussion on the underlying structures and assumptions. Finally, limitations of the study and recommendations for future research are given.

2 Literature Review

Debates on ethical challenges arise especially in times of significant technical progress. Beginning with early developments of computers, Wiener stated concerns regarding automated work forces and application in warfare in the 1940s and 1950s (Bynum 2010). In the 1960s, Parker and Maner called attention to the increasing crimes committed with the help of computers and by technicists (Bynum 2010). Moor (1985) provides an explanation for the numerous ethical problems caused by computers. The author characterizes computers as logically malleable which appears limitless in the potential application. For activities that have not been able without computers people cannot rely on good practices or ethical standards, a state of “policy vacuum” as Moor (1985) calls it. The author sees the role of computer ethics in dissolving the conceptual muddle by ethical reflection of concrete cases and providing a “coherent conceptual framework within which to formulate a policy for action” (Moor 1985, 266). Since 1985, the level of abstraction has shifted from technological means to information as content to data as the smallest entity. The focus followed the technical development and drew the attention to the points where ethical problems are likely to arise (Floridi and Taddeo 2016, 3). Still, Moor’s understanding of the mandate remains valid nowadays. It can be traced in the definition of Data Ethics proposed by Floridi and Taddeo (2016) which determines it as “the branch of ethics that studies and evaluates moral problems related to data […], algorithms […] and corresponding practices […], in order to formulate and support morally good solutions (e.g. right conduct or right values)” (3). There seems to be deficit consensus regarding the denomination as the term “Data Ethics” is one among others – including the authors of the same definition making use of expressions like “AI ethics” in more recent articles (e.g. Morley et al. 2020).

Albeit the terminology remains ambiguous, Moor’s understanding of the discipline’s contribution serves to classify the existing research. The existing research on the one hand studies ethical challenges and on the other hand fathoms suggestions for right conduct. Ethical deliberation in CIE is often drawing on normative ethics, a branch “that is concerned with establishing how things should or ought to be, how to value them, which things are good or bad, and which actions are right or wrong” (Dignum 2017, 3). Virtue ethics, deontology, and consequentialism are the normative theories that are most prominently referred to, focusing on the morality of the acting person, the act itself or the outcome (Dignum 2017; Kraemer, van Overveld, and Peterson 2011; Mittelstadt et al. 2016; Saltz and Dewar 2019; Sandvig et al. 2016). A loose interpretation of all three normative theories turns out most promising in practice for guiding ethical assessment in CIE, as Ananny (2016) and Sandvig et al. (2016) recommend.

2.1 Ethical Challenges and Right Conduct

Several authors have considered the impact of personal beliefs and preferences during the design and development process. Stereotypes, values, and worldviews held by the persons involved in the development affect data collection, algorithm building, and choice of certain models and
focus on the biased suspects that systemic discrimination is upheld by based on those results (Mittelstadt et al. 2016). Hoffman groups and individuals, especially if decisions are taken to reinforce discrimination of marginalized and vulnerable baum 1996; Introna 2005; Kraemer, van Overveld, and Peterson 2011). The outcomes of biased algorithms are likely to reinforce discrimination of marginalized and vulnerable groups and individuals, especially if decisions are taken based on those results (Mittelstadt et al. 2016). Hoffman (2019) suspects that systemic discrimination is upheld by ignoring the underlying social problems and criticizes the focus on the biased “bad actors” at the expense of shared responsibility. Disrespect of principles like privacy, autonomy, and beneficence may also cause ethical problems. Effectively ensuring anonymity becomes difficult as data gets aggregated (Saltz and Dewar 2019; Zwitter 2014) and for data subjects impedes understanding if their privacy consent is complied with (Ananny 2016; Mittelstadt et al. 2016). Deficient traceability complicates autonomous operating and effectively determining one’s identity, e.g. by nudging user behavior (Richards and King 2014). Responsibility for possible harmful incidents is another crucial aspect. Various researchers attempted to trace back contributions to individual persons (Ananny 2016; Dignum 2017; Mittelstadt et al. 2016). Since this may lead to shifting responsibility between actors (Taylor 2017), Leonelli (2016) suggests shared accountability by all involved persons.

Early propositions to address ethical dilemmas with computers were awareness raising by installing codes of conduct and providing education for engineers. In the 1970s, Parker advanced the first Code of Professional Conduct for the Association of Computing Machinery (ACM) and Maner realized an experimental course on computer ethics and teaching material to advise students of computer science (Bynum 2010). Both approaches remain popular to address upcoming challenges. Saltz and Dewar (2019) view the assessment process as encouraging for critical thinking whereas Leonelli (2016) worries about an “outsourcing” of ethical concern. A multitude of codes, guidelines, and frameworks has been published and studied quantitatively (Hagendorff 2020; Jobin, lenca, and Vayena 2019; Schiff et al. 2020; Whittlestone et al. 2019b). Teaching ethics in computer science and data science has not evolved to a standard but increasingly forms part of curriculums, exploring adequate instruction methods (Celis 2019; Shapiro et al. 2020). Another suggestion is disclosure of choices and assumptions during the technical design process that allow to judge the context (Gebru et al. 2021; Kraemer, van Overveld, and Peterson 2011; Steen 2015). Similar to that, many studies – especially from engineering disciplines – have explored technical means to avoid discrimination or respect values like privacy (Dunkelau and Leuschel 2019; van den Hoven, Vermaas, and van de Poel 2015).

2.2 Human Values

In CIE, a common thread has been “the concern for protection and advancement of major human values” (Bynum 2010, 34). Values are understood as the morally ideal human behavior on an abstract, societal level “to promote the right course of action” (Brey 2010, 47). Wiener and Moor attempted to determine core values (Bynum 2010), whereas recently researchers developed amplified value sets (La Fors, Custers, and Keymolen 2019). This evolution illustrates an unresolved tension between the aim of unification (Floridi 2019) and the acknowledgment of complex and singular situations which require specific principles (van den Hoven 2010; Vayena and Tasioulas 2016). Furthermore, the interplay between technology and values is reciprocal: human values shape the development of technology as well as technology may shape the values held by humans (Nissenbaum 2001; Richards and King 2014). The Value Sensitive Design (VSD) approach derives an interactional stance from those intertwined relations: developers and designers are assumed to have room for consciously endorsing values by embodying them into devices and systems (Friedman and Hendry 2019). This requires visibility and explication of the promoted values to support comprehensibility and common ground.

VSD releases an extensive methodology toolkit for determination of values at stake and analyses of direct and indirect stakeholders, among others (Friedman and Hendry 2019). The latter is essential to recognize deviant values and interests between involved persons and groups. This is classified as the epistemological source of value conflicts in contrast to the ontological source of conflicts raised by trade-offs between various values (Manders-Huits 2011). Whittlestone et al. (2019a) claim to focus on tensions between values since this reveals different interpretations, requirements for new solutions, and knowledge gaps and thus more fruitfully guides conduct than general principles. In concordance with Friedman and Hendry (2019), the authors propose “extensive public engagement” (Whittlestone et al. 2019a, 199) to understand the respective needs and values. Yet, applicable methodologies to weigh interests and values are largely unexplored.

Overall, current studies support addressing concrete ethical problems and attempt to formulate guiding principles although no consensus is reached. Awareness raising and providing instructions for action are the predominant approaches to solving problems, and increasingly technical means are developed. Research gaps become apparent at handling value conflicts or the weighing of values and at translating ethical principles into practical work.
3 Methodology

The acknowledgment of the reciprocal influence of technology and values indicates a constructivist notion. Consequently, a constructivist procedure, namely Critical Discourse Analysis (CDA), is employed to examine four Data Ethics Frameworks. The specificities of the qualitative method will be presented as well as sampling of the frameworks, analysis of language patterns, and creation and publication of texts.

Social constructivism can be described as the idea “that our knowledge of the world, including our understanding of human beings, is a product of human thought rather than grounded in an observable, external reality” (Burr 2015, 222). As persons we are formed by the culture, norms, and situations surrounding us. Language as a key aspect of social interaction is therefore essential to social constructionism and results in a number of theories and approaches centering on discourse (Burr 2015). One of those is CDA by Fairclough (1995), a well-established concept that supports studying power structures in language and takes an interactional perspective towards meaningful social change.

CDA approaches are generally characterized by their assumption of ideological language, the constructive relation between language and social practice, as well as the critical view from the perspective of the oppressed group (Jørgensen and Phillips 2002). CDA as coined by Fairclough builds theoretically on Marxist academics for his understanding of ideology and hegemony. Ideology not only is perceivable in language but also in social practices and becomes invisible through the recognition as common-sense (Fairclough 1995). The competition of the prevailing social class surfaces at the level of discursive practice i.e., the activities in which language and texts are embedded. This so-called hegemonic struggle “contributes in varying degrees to the reproduction or transformation of the existing order of discourse, and through that of existing social and power relations” (Fairclough 1995, 77). Thus, people have a range of possibilities to act, to use language creatively, change meanings, and to resist.

Fairclough’s CDA framework considers three dimensions of discourse corresponding to different discourse analysis techniques (Figure 1). Text is at the core as it is embedded into certain practices of text production, dissemination, and interpretation (discourse practice) and constructed by customs, beliefs, and conduct in a specific social field (sociocultural practice). To detect the power relations and ideological strains in language, a linguistic analysis is conducted at text level. Discourse practice is addressed by a critical review of the established processes of text production. The inclusion of social theory permits deducing explanations of the consequences of reproduction or change for the wider sociocultural practice (Jørgensen and Phillips 2002).

3.1 Sampling

Fairclough does not specify clearly how to select the samples – his own exemplified analyses consist of single texts, advertisements or phrases rather than a corpus (Fairclough 2001). Jørgensen and Phillips (2002) recommend a sample that adequately supports the assumptions. Thus, four comparable frameworks were retrieved from the AI Ethics Global Inventory\(^1\) run by the German watchdog organization DE GRUYTER.

---

\(^{1}\) See https://inventory.algorithmmwatch.org/, accessed June 29, 2021.
AlgorithmWatch: Data Ethics Decision Aid (DEDA) (Schäfer and Franzke 2020), Digital Decisions Tool (DD Tool) (Center for Democracy & Technology 2017), Data Ethics Canvas (DEC) (Open Data Institute 2019a), and Data Ethics Workbook (DEW) (Department for Digital, Culture, Media & Sport UK 2018). They are similar in their scope, their questionnaire (DEW) (Department for Digital, Culture, Media & Sport UK 2018).3 They are similar in their scope, their questionnaire genre, and their appealing and handy design. Table 1 gives an extensive overview of the features. All frameworks are released by public or non-profit organizations between 2016 and 2017 and for comparable target groups and aims. The greatest difference consists in how engaged organizations continue the work with updates, support options and supplementary material, and how they pursue dissemination of their frameworks.

3.2 Methods of Analysis

Text analysis is carried out by assessing clause combination, modality, vocabulary usage, and cohesion (Fairclough 1989). This allows one to detect how aspects of the world and persons are represented and connected (Fairclough 2001). Moreover, the conception of discourses implies how common language modes are used or whether new meanings are assigned to the words. As opposed to previous studies, values are derived from the integrated value list by La Fors, Custers, and Keymolen (2019) and tested for their consideration. The listed values are human welfare, autonomy, non-maleficence, justice (incl. equality, non-discrimination, digital inclusion), accountability (incl. transparency), trustworthiness (incl. honesty and security), privacy, dignity, solidarity, and environmental welfare (La Fors, Custers, and Keymolen 2019, 214). In preparation for text analysis, transcripts of the plain text were made. In a first cycle textual patterns were identified at the level of grammar, vocabulary, and clause construction. A second cycle connected those with values, discourses, and assigned coded elements to designers and stakeholders. Process analysis draws on the websites of the organizations who issued the frameworks. The texts and blog articles comment on the process of creation, describe how it was disseminated and adopted, and thereby reveal established discursive practices. The results of analysis are discussed against the background of beliefs, attitudes, and structures present in CIE. In line with social constructionism, most relevant are the processes around knowledge creation, “the taken-for-granted ways of understanding the world” (Burr 2015, 223). Emphasis is laid on the suspected common-sense claims that are not challenged as they represent the dominant view. The literature review supplies the basis and is backed by recent critical approaches like Critical Data Studies and feminist and post-colonial theories. Validity, generalization, and verification are not agreed upon in CDA and discourse analysis in general (Jørgensen and Phillips 2002), although propositions exist like triangulation with other methods or diverse material and exhaustive analysis (Meyer 2001). Coherent analysis and transparent discussion of inconsistencies as well as disclosure of personal attitudes towards the subject are thus applied to allow comprehension and traceability from other researchers.

4 Results

In this section the identified values and value conflicts are reported and illustrated with the help of brief examples. Furthermore, the prevailing discourses and power structures are assessed and complemented with discursive practices of the organizations.

In total, 646 codes were assigned to the four frameworks. Of this sum, 194 codes were distributed to values and value conflicts. Evidence for all values of the integrated list was found except for environmental welfare. However, dignity, trustworthiness, and solidarity were not referred to under those denominations which made them difficult to distinguish from other values and thus were merged. On the contrary, transparency was singled out as a separate value for its prominent appearance. Various discourses had been detected due to corresponding vocabulary: a business, legal, technological, anti-discrimination, democratization, and general value discourse (246 codes). The remaining 233 codes were assigned to the “designers” – the target group of the frameworks – and the “stakeholders” who are mentioned as objects to technologies and devices. Power structures manifest in the presentation of this group of persons. Table 2 gives insights on the quantitative distributions of the codes among DEFs.

5 Extracted Values and Discourses

Human welfare is often referred to as “benefit” and “user need” which should guide the development process and is present in three of four frameworks. Present tense indicates certainty of the additional value (“What are the benefits of the project?”, DEDA, l. 12). Emphasis is placed on the

2 Analyzed version: May 2019.
3 Analyzed version: September 2018.
4 Transcripts, code book and self-reflection are published at OSF: https://osf.io/z7e9m/?view_only=b4e299c37ab84/55b28803786f9df88f.


| Feature                      | Data Ethics Decision Aid (DEDA) | Digital Decisions Tool (DD Tool) | Data Ethics Canvas (DEC) | Data Ethics Workbook (DEW) |
|------------------------------|---------------------------------|---------------------------------|-------------------------|---------------------------|
| Organization                 | Utrecht Data School (University Utrecht) | Center for Democracy and Technology (CDT) | Open Data Institute (ODI) | Central Digital & Data Office (CDDO)/Cabinet Office |
| Legal form                   | Public                          | Non-profit                      | Non-profit              | Public                    |
| Size (headcount)             | 17 (Utrecht Data School 2021b)  | 29 (Center for Democracy & Technology 2021) | 64 (Open Data Institute 2021a) | Not known                  |
| Country                      | Netherlands                     | US/Belgium                      | United Kingdom          | United Kingdom            |
| Creation date                | End of 2016/beginning of 2017  | August of 2017                  | September of 2017       | May of 2016               |
| Last update                  | June of 2020                    | Not known                       | June of 2021            | September of 2020         |
| Purpose                      | To help recognize ethical issues; foster accountability, educate users, communicate problems, support project management (Utrecht Data School 2020a) | To help “understand and mitigate unintended bias and ethical pitfalls” in the design of automated decision-making systems (Duarte 2017) | To help “identify and manage ethical issues” (Tarrant and Maddison 2021) | To help “understand ethical considerations, address these within […] projects, and encourage[s] responsible innovation” (Government Digital Service 2020) |
| Target group                 | Data analysts, project managers, policymakers, interdisciplinary teams, stakeholders (Utrecht Data School 2020a) | Engineers and product managers, design level, developers (Duarte 2017) | “[A]nyone who collects, shares or uses data” (Open Data Institute 2019a) | “[A]nyone working directly or indirectly with data in the public sector, including data practitioners (statisticians, analysts and data scientists), policymakers, operational staff and those helping produce data-informed insight” (Central Digital & Data Office 2020) |
| Creation process             | Developed by students at Utrecht Data School, in cooperation with data analysts from City of Utrecht (Utrecht Data School 2020b) | Created by CDT and “others”; “informed by extensive research” (Duarte 2017) | Supposedly created by ODI; based on the Ethics Canvas by ADAPT Center for Digital Content Technology and the Business Model Canvas by Alex Osterwalder (Tarrant and Maddison 2021) | Supposedly developed by the cabinet office team; revisions included public and expert engagement (Cabinet Office 2016; Central Digital & Data Office 2020) |
| Publication & dissemination  | Website, AI Ethics Guidelines Global Inventory, City of Utrecht, The Association of Dutch Municipalities, academic publications (Franzke, Muis, and Schäfer 2021) | Website, AI Ethics Guidelines Global Inventory | Website, AI Ethics Guidelines Global Inventory, ODI Community (Vryzakis and Thereaux 2020) | Website, AI Ethics Guidelines Global Inventory, Municipalities |
| Supplementary material       | Handbook with brief explanations and examples for the questions (Utrecht Data School 2020b) | Not available                   | White paper with review of existing practices in organizations; user guide (not openly accessible) (Tarrant and Maddison 2021) | Glossary, links for legislation and codes of practice for use of data (Central Digital & Data Office 2020) |
| Support                      | Optional introductory workshops for teams (paid) (Utrecht Data School 2020c) | Not available                   | Introductory workshops and Data Ethics courses (paid) (Open Data Institute 2021b) | Not known                  |
| License                      | All rights reserved             | CC BY                           | CC BY SA 4.0            | Open Government License   |
communication and enhancement of positive impact which follows a business logic (“How are you measuring and communicating positive impact? How could you increase it?”, DEC, l. 40–41). Benefit is therefore juxtaposed with business advantage and financial revenue. In discursive practice, this corresponds with the positive storytelling communication especially applied by the Open Data Institute (2019b).

In contrast to human welfare, non-maleficence is determined more specifically since causes of harm are supposedly more familiar. Bias, misuse, and misinterpretation are prominently mentioned but use of modal verbs and passive forms intend to create a distance to the project and disguise responsible actors (“What are the problems or concerns that might arise in connection with this project?”, DEDA, l. 13). The notions of misuse and misinterpretation imply the ambition to uphold authority of “right” use and interpretation. Harmful incidents seem to be a potential business risk also in terms of public criticism (“Does the project risk generating public concern or outrage?”, DEDA, l. 43). Non-maleficence also includes knowing one’s limits and consulting external experts. That aspect is well covered in DEW although expertise appears to be closely related with formal education (“subject matter experts”, DEW, l. 36). Finally, a precaution principle is observed among all frameworks to consider possible long-term implications.

Justice is regarded in terms of fair and equal treatment and freedom of discrimination and thereby underpins dignity. DEDA refers to justice and inclusion as values, while other frameworks mention eradication of bias at the level of data, algorithm, and outcomes (“Where could bias have come into this analysis?”, DD Tool, l. 92). The broad coverage of bias mirrors the extensive public and academic debate and relates to further activities by the organizations.

The Center for Democracy and Technology for instance carried out a project in that domain whose results directly inspired DD Tool (Lange 2016) and Utrecht Data School employ a new project under the acronym BIAS (Utrecht Data School 2021a). The questions are on the one side reinforcing a “bad actor frame” (Hoffman 2019, 903) by locating the source of bias in the assumptions of single persons or homogeneous teams and on the other side transport a tendency of technological determinism that humans are surrendered to (Greene, Hoffman, and Stark 2019). Solutions for achieving justice are framed by a technology discourse as optimizing technologies are considered (“Did your feedback mechanism capture and report anomalous results in a way that allows you to check for biased outcomes?”, DD Tool, l. 84). An anti-discrimination discourse is perceptible when people are given the opportunity to share their experience and are taken seriously (“Do citizens have the opportunity to raise objections to the results of the project?”, DEDA, l. 45).

La Fors, Custers, and Keymolen (2019) see transparency as backing accountability but the frameworks recognize transparency separately. It is understood in terms of publishing openly and communicating understandably (“Could you publish your methodology, metadata, datasets, code or impact measurements?”, DEC, l. 60). Therefore, it is related to a business discourse making use of strategic communication (“What is the communication strategy with regard to this

Table 2: Quantitative presentation of codes for each framework.

| Code                      | DEDA (n = 128) | DD tool (n = 192) | DEC (n = 171) | DEW (n = 155) |
|---------------------------|---------------|------------------|--------------|---------------|
| Human welfare (& solidarity) | 1 (2)         | 0 (0)            | 3 (3)        | 2 (2)         |
| Autonomy (& trustworthiness)   | 4 (6)        | 10 (5)           | 7 (3)        | 2 (3)         |
| Non-maleficence            | 4             | 4                | 2            | 3             |
| Justice (& dignity)        | 6             | 16 (3)           | 5 (3)        | 2 (3)         |
| Accountability             | 4             | 7                | 5            | 11            |
| Transparency               | 5             | 4                | 11           | 4             |
| Privacy                   | 8             | 4                | 3            | 5             |
| Value conflicts            | 7             | 3                | 1            | 7             |
| Value discourse           | 2             | 2                | 1            | 0             |
| Business discourse        | 15            | 11               | 20           | 14            |
| Legal discourse           | 6             | 2                | 6            | 6             |
| Technology discourse      | 7             | 27               | 4            | 10            |
| Anti-discrimination discourse | 8           | 28               | 16           | 4             |
| Democratization discourse | 2             | 7                | 10           | 11            |
| Designer                  | 29            | 29               | 38           | 40            |
| Stakeholder               | 11            | 30               | 30           | 26            |
A legal discourse comes into play as certain duties are exemplified (“Are non-deterministic outcomes acceptable given your legal or ethical obligations around transparency and explainability?”, DD Tool, l. 52). Organizations themselves handle openness differently: Utrecht Data School and Central Digital & Data Office publish their frameworks and supplementary material freely accessible (newest versions) whereas some reports of Center for Democracy & Technology needed to be retrieved from an internet archive. Open Data Institute, the one that emphasized transparency and openness, offers a free download of the framework, but the user guide and other publications are accessible from a commercial platform only after registration. Furthermore, organizations are often reluctant at openly disclosing their motives, contributors, and understanding of Data Ethics.

Accountability is used interchangeably with responsibility and is aimed at ensuring traceability. This results in a distribution of responsibility for ethical challenges to individuals (“Is there a person on your team tasked specifically with identifying and resolving bias and discrimination issues?”, DD Tool, l. 9) or among organizational hierarchies (“How often will you report on these plans to senior reporting officers?”, DEW, l. 58). A considerable legal discourse demonstrates that responsibility is often interpreted with regards to existing liabilities (“Which laws and regulations apply to your project?”, DEDA, l. 35). Relation with transparency indicates documentation or preparation for audit. The organizations that released the frameworks refuse to take any accountability for outcomes and implications of the ethical deliberation (Broad, Smith, and Wells 2017; Utrecht Data School 2020b).

Autonomy, the ability to pursuing own thoughts, will, goals, and decision-making, starts at the very beginning of being involved into a certain project or data collection (“Was the data collected in an environment where data subjects had meaningful choices?”, DD Tool, l. 28). Furthermore, it is about the means of interaction within the project and with the creators of a technology or device. Untargeted collaboration is mentioned in two frameworks (“Are you routinely building in thoughts, ideas and considerations of people affected in your project? How?”, DEC, l. 65). In many phrases, stakeholders are referred to either pertaining to a passive collective or according to their sensitive features. While the feelings and thoughts of the designers are given room, stakeholders are not deemed the same position to utter feelings apart from discrimination. In practice of framework creation, the organizations headed their development, but acquired support from the practice (Utrecht Data School 2020a) or at least conducted user studies for revision (Central Digital & Data Office 2020; Ginnis et al. 2016). Caring about instruction and adjusting it to different groups strengthens democratization since knowing an issue supports forming of opinions and acting autonomously (“What information or training might be needed to help people understand data issues?”, DEC, l. 66).

Privacy is understood as sensitive personal information and predominantly addressed in the frame of a legal discourse (“If using personal data, do you understand obligations under data protection legislation?”, DEW, l. 13). The European General Data Protection Regulation (GDPR) is strict in that sense to protect individuals and work towards data minimization. Additionally, it is referred to other elements of GDPR and shows the influence of regulations even in an ethical context that could go further than the requirements of law (“Have you conducted a PIA (Privacy Impact assessment [sic]) or DPIA (Data Protection Impact Assessment)?”, DEDA, l. 49). Since present tense indicates the normalization of processing of personal data, data minimization is an aspect that entails an alternative (“How can you meet the project aim using the minimum personal data possible?”, DEW, l. 18). Moreover, anonymization techniques, access control mechanisms, and synthetic data are listed as options and illustrate the technical discourse applied to that value.

DEDA is the only framework that encourages reflecting on personal and organizational values and considers that it may give inconsistencies and conflicts. This issue was adjusted due to experiences in practice (Franzke, Muis, and Schäfer 2021). Other frameworks vaguely refer to conflicts of interest between project aims (“Are you replacing another product or service as a result of this project?”, DEC, l. 33) and stakeholder groups (“Is there a fair balance between the rights of individuals and the interests of the community?”, DEW, l. 24). However, little guidance is given to examine how “fair” might be interpreted and how the various interests are documented. Values may come into conflict where supposed project benefits interfere with individuals’ privacy or autonomy or among the project team (“Are all parties involved in agreement as to this strategy?”, DEDA, l. 40).

Overall, these results show that the identified values of human welfare, non-maleficence, justice, transparency, accountability, autonomy, and privacy are addressed in all frameworks. The principles are often listed singularly which strengthens the impression that value conflicts are omitted. Surprisingly, the general value discourse is marginal compared to the business, technology, legal, anti-discrimination, and democratization discourse. At the level of discursive practice, the issuing organizations headed framework creation although various stakeholders were
often included at later stages of the process. Particularly in terms of transparency, many organizations do not comply with their own ambitions transported in the DEFs. Apparently, due to the focus on the perspective of their target group, anyone who deals with data, it is out of sight that stakeholders and data subjects may support certain values as well.

6 Discussion

The findings of this analysis show similarities with other studies that examined ethical guidelines (Hagendorff 2020; Jobin, Ienca, and Vayena 2019; Schiff et al. 2021). Values like transparency, privacy, accountability, and justice are predominantly mentioned across the four frameworks and likewise are the most covered ones in the mentioned studies. On the contrary, values are not opposed among each other nor are conflicts between groups or interests amplified. The academic debates in CIE are reflected by the same importance of topics as principles are better researched than value conflicts. This connection demonstrates the mutual construction of discourse and sociocultural practice. In the tradition of social constructionism and CDA, implications for knowledge creation and power structures are discussed in the following.

6.1 What is Regarded as Knowledge?

Generally, data processing and the application of algorithms are normalized across the frameworks. This technopositivist view is not challenged by the question of whether data science is always the appropriate solution to address a problem. In CIE, most researchers circumvent the ambiguity and take the dominant view of using data science (Floridi and Taddeo 2016). This common sense narrows the option to argue that algorithmic systems are not always a good solution (Greene, Hoffman, and Stark 2019; Powles and Nissenbaum 2018). It is therefore noteworthy that in June of 2020 IBM decided to suspend the distribution of facial recognition systems with reference to the values in the company’s ethical guidelines (Krishna 2020).

The DEFs regard knowledge as conceived from the available data. Questions about data collection methodologies challenge the circumstances of data collection, but it is not generally disputed that adequate data exist for application in the project. This assumption neglects the (often immaterial) labor related to data production and generation (Amrute 2019; Fotopoulou 2019). Disregarding those reflections and processes not only ignores the sensitivity for discrimination, but more pressing is that it follows that complex reality can be adequately represented in data. Partial and contingent forms of knowledge, abilities, and cultural wisdom might not be possible to transform into binary code. Consequently, those aspects of people’s reality are not incorporated and become invisible.

6.2 Who Participates in the Discourse?

In terms of the actors participating in the ethical debate, the interdisciplinary teams of organizations like Utrecht Data School prove how the field has come away from a hegemony of computer scientists and technical skills (Boyd and Crawford 2012). Other disciplines and roles apart from programming are deemed relevant as the prominent communicative aspects and business discourse indicate. However, little has changed in the way how a small group – the developers, project managers and designers – determines how technology is used and “who gets to participate” (Boyd and Crawford 2012, 675). The deficient methodology for stakeholder collaboration in CIE is a proof of the lacking practice (Manders-Huits and Zimmer 2009). Probably other stakeholders are suspected not to contribute valuable input. This gap could be closed by accounting for the contingent and situated knowledge held both by researchers and other stakeholders and illustrating the same situated context in which “knower and known” operate (Croupe and Linabary 2020, 156). Disclosing the researchers’ contingency towards methods and subjects is not established yet in academia but regarded as a fruitful and applicable way to a reflective research (Croupe and Linabary 2020).

Who gets to participate is especially relevant with regards to anti-discrimination. The ones that are vulnerable to discrimination through biased algorithms or data collections are often those that do “not […] arrive in the present with equal power or privilege” (D’Ignazio and Klein 2020, 152). As the project affiliates like designers and project managers determine the degree of participation, it suffices with complaints and loose feedback and “serves as a mere legitimation exercise” (Schiff et al. 2021, 40). The dominant view becomes apparent with regard to opponent concepts. D’Ignazio and Klein (2020) for instance propose Data Justice, an approach with emphasis on reparative justice to respect prior inequity. Ethical data is deemed not sufficient as the authors of the Good Data Manifesto write and claim for data that is actively pursuing “good” (Trenham and Steer 2019). Inclusion of stakeholders during the creation process would disrupt common forms of text creation and is suspected to diversify terms, how they are conceptualized and the meaning that is ascribed to the
language. In the CIE discipline, there is an imbalance in favor of Global North dominance as reported by Jobin, Ienca, and Vayena (2019). Recently, Data Ethics initiatives are initiated in countries of the Global South as a workshop at the 2021 ACM Web Science Conference illustrates. The controversy around the cancellation of Timnit Gebru, AI ethics researcher at Google, raised questions to what extent people who utter internal criticism against practices and who advocate in AI research as black woman are desired to form part of the debate (Simonite 2021).

The ambiguous relation between CIE and tech industry is expressed well with the hegemonic struggle around the identified discourses. Values are not only related to an ethics discourse but also in terms of business. Practical implications show how some organizations make financial revenue with their courses on Data Ethics. It is a necessary objective to gain acceptance for the consideration of ethical deliberation within the tech industry – an ambitious goal as Mittelstadt (2019) states and a challenging task for corporate ethicists (Metcalf, Moss, and Boyd 2019). Yet, it should be questioned how this dependency affects the academic discussions in CIE. Close entanglement becomes obvious when Facebook funds an Institute for Ethics in Artificial Intelligence at Technical University of Munich (Köver and Dachwitz 2019) or biased research conducted by MIT for the benefit of Silicon valley industry is witnessed (Ochigame 2019). In line with CDA objectives, it should be discussed how the academic debate can be fostered without dilution in industry. Mittelstadt (2019) for instance comments on the necessity for high-level theories which are able to translate into requirements in practice. Social change can thus be observed not in the sense of improving the situation for the oppressed as it is intended in CDA, but the present and powerful actors are rather strengthened.

6.3 Limitations

This study is limited in its generalization as coding and interpretation was carried out by one researcher. Even though it is attempted to ensure comprehension and traceability as recommended by qualitative literature, external validity could be increased by inter-rater reliability coding. As a means of triangulation, previous versions of the respective frameworks could be included since several editions exist for most of them. The deductive method of coding with help of the integrated value list by La Fors, Custers, and Keymolen (2019) proved to be applicable. However, definitions of the principles were not facilitated by the authors which made it in some cases difficult to distinguish values.

7 Conclusions

In this study, four practically designed Data Ethics Frameworks from public or non-profit institutions were investigated to identify the promoted values and evaluate the representation of value conflicts. Findings show a set of established values which is present in all frameworks although emphasis differs across the publications. This indicates a close relation with the information ethics discipline which has been occupied from the beginning with preserving human values. Although language structures and values indicate a reinforcement of established practices and customs, a hegemonic struggle between various actors can be observed. Values are increasingly interpreted as a business factor and thus related to aspects of communication, legal compliance, and technological solutions. Concerns on eradicating discrimination allow for an anti-discrimination and democratization discourse but reinforced power asymmetries weaken the effectiveness. Since the frameworks take the perspective of their target group, affected data subjects are contemplated from a distance and not meaningfully included into the debate, neither via DEFs at text level nor in terms of discursive practices in the moment of text creation. Therefore, it is recommended to apply and test means of participation of diverse direct and indirect stakeholders because it is still unexplored in research. Focusing on intersection of values and value conflicts should also play a greater role in research. This has the potential to advance the academic and public debate in the question of which values should be prioritized, and which trade-offs might be acceptable at designing future technologies.

References

Amrute, S. 2019. “Of Techno-Ethics and Techno-Affects.” Feminist Review 123: 56–73, https://doi.org/10.1177/0141778919879744.

Ananny, M. 2016. “Toward an Ethics of Algorithms.” Science, Technology & Human Values 41 (1): 1–25, https://doi.org/10.1177/0162243915606523.

Boyd, D., and K. Crawford. 2012. “Critical Questions for Big Data.” Information, Communication & Society 15 (5): 662–79, https://doi.org/10.1080/1369118X.2012.678878.

Brey, P. 2010. “Values in Technology and Disclosive Computer Ethics.” In The Cambridge Handbook of Information and Computer Ethics, edited by L. Floridi, 41–58. Cambridge: Cambridge University Press.
Fairclough, N. 2016. “Long-Term Trends in the Public Perception of Artificial Intelligence.” https://arxiv.org/abs/1609.04904 (accessed June 29, 2021).

Floridi, L. 2019. “Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical.” Philosophy & Technology 32 (2): 185–93. https://doi.org/10.1007/s13347-019-00354-x.

Floridi, L., and M. Taddeo. 2016. “What Is Data Ethics?” Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences 374 (2083). https://doi.org/10.1098/rsta.2016.0360.

Fotopoulou, A. 2019. “Understanding Citizen Data Practices from a Feminist Perspective: Embodiment and the Ethics of Care.” In Citizen Media and Practice: Currents, Connections, Challenges, 1st ed., edited by H. C. Stephansen, and E. Trré. London: Routledge. Author’s submitted copy - prepublication copy.

Franze, A. I., I. Muis, and M. T. Schäfer. 2021. “Data Ethics Decision Aid (DEDA): A Dialogical Framework for Ethical Inquiry of AI and Data Projects in the Netherlands.” Ethics and Information Technology, https://doi.org/10.1007/s10676-020-09577-5.

Friedman, B., and D. G. Hendry. 2019. Value Sensitive Design: Shaping Technology with Moral Imagination. Cambridge: MIT University Press.

Friedman, B., and H. Nissenbaum. 1996. “Bias in Computer Systems.” ACM Transactions on Information Systems 14 (3): 330–47.

Gebru, T., J. Morgenstern, B. Vecchione, J. Wortman Vaughan, H. Wallach, H. Daumé III, and K. Crawford. 2021. “Datasheets for Datasets.” Communications of the ACM 64 (12): 86–92.

Ginnis, S., H. Evans, N. Boal, E. Davies, and A. P. Aslaksen. 2016. “Public Dialogue on the Ethics of Data Science in Government.” https://www.ipsos.com/sites/default/files/2017-05/data-science-ethics-in-government.pdf (accessed June 29, 2021).

Government Digital Service. 2020. “Data Ethics Framework.” https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/923108/Data_Ethics_Framework_2020.pdf (accessed June 29, 2021).

Greene, D., A. L. Hoffman, L. Stark. 2019. “Better, Nicer, Clearer, Fairer: A Critical Assessment of the Movement for Ethical Artificial Intelligence and Machine Learning.” In Hawaii International Conference on System Sciences (HICSS). Maui, Hi. http://dmgreene.net/wp-content/uploads/2018/09/Greene-Hoffman-Stark-Better-Nicer-Clearer-Fairer-HICSS-Final-Submission.pdf (accessed June 29, 2021).

Haas, L., and S. Gießler. 2020. “In the Realm of Paper Tigers – Exploring the Failings of AI Ethics Guidelines.” https://algorithmwatch.org/en/ai-ethics-guidelines-inventory-upgrade-2020/ (accessed June 29, 2021).

Hagendorff, T. 2020. “The Ethics of AI Ethics: An Evaluation of Guidelines.” Minds and Machines 30 (1): 99–120. https://arxiv.org/ftp/arxiv/papers/1903/1903.03425.pdf (accessed June 29, 2021).

Hoffman, A. L. 2019. “Where Fairness Fails: Data, Algorithms, and the Limits of Antidiscrimination Discourse.” Information, Communication & Society 22 (7): 900–15. https://doi.org/10.1080/1369118X.2019.1573912.

Introna, L. D. 2005. “Disclosive Ethics and Information Technology: Disclosing Facial Recognition Systems.” Ethics and Information Technology 7 (2): 75–86. https://doi.org/10.1007/s10676-005-4583-2.
Taylor, L. 2017. “What Is Data Justice? The Case for Connecting Digital Rights and Freedoms Globally.” Big Data & Society 2: 1–14, https://doi.org/10.1098/rsta.2016.0126.

Trenham, C., and A. Steer. 2019. “The Good Data Manifesto.” In Good Data, Theory on Demand 29, edited by A. Daly, S. K. Devitt, and M. Mann, 37–53. Amsterdam: Institute of Network Cultures.

Utrecht Data School. 2020a. “Data Ethics Decision Aid (DEDA).” https://dataschool.nl/deda/?lang=en (accessed June 29, 2021).

Utrecht Data School. 2020b. “Handbook: Assessing Ethical Issues with Regard to Governmental Data Projects.” https://dataschool.nl/wp-content/uploads/sites/272/2020/06/DEDA-Handbook-ENG-V3.1-1.pdf (accessed June 29, 2021).

Utrecht Data School. 2020c. “Workshop.” https://dataschool.nl/deda/workshop/?lang=en (accessed June 29, 2021).

Utrecht Data School. 2021a. “Beraadslagingsinstrument Voor Algoritmische Systemen (BIAS).” https://dataschool.nl/en/samenwerken/bias/ (accessed June 29, 2021).

Utrecht Data School. 2021b. “Team.” https://dataschool.nl/en/about-uds/team/ (accessed June 29, 2021).

van den Hoven, J. 2010. “The Use of Normative Theories in Computer Ethics.” In The Cambridge Handbook of Information and Computer Ethics, edited by L. Floridi, 59–76. Cambridge: Cambridge University Press.

van den Hoven, J., P. E. Vermaas, and I. van de Poel, eds. 2015. Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains. Dordrecht: Springer.

Vayena, E., and J. Tasioulas. 2016. “The Dynamics of Big Data and Human Rights: The Case of Scientific Research.” Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences 374 (2083), https://doi.org/10.1098/rsta.2016.0129.

Vryzakis, A., and O. Thereaux. 2020. “How Our Network Is Considering Data Ethics: Survey Results.” https://theodi.org/article/how-our-network-is-considering-data-ethics-survey-results/ (accessed June 29, 2021).

Whittlestone, J., R. Nyrup, A. Alexandrova, and S. Cave. 2019a. “The Role and Limits of Principles in AI Ethics: Towards a Focus on Tensions.” In Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, edited by AIES ’19, 195–200. New York: Association for Computing Machinery (accessed June 4, 2020).

Whittlestone, J., R. Nyrup, A. Alexandrova, K. Dihal, and S. Cave. 2019b. Ethical and Societal Implications of Algorithms, Data, and Artificial Intelligence: A Roadmap for Research. London: Nuffield Foundation.

Zwitter, A. 2014. “Big Data Ethics.” Big Data & Society 1 (2): 1–6, https://doi.org/10.1177/2053951714559253.