

# AI Ethics in the Public, Private, and NGO Sectors: A Review of a Global Document Collection

Daniel Schiff, *Graduate Student Member, IEEE*, Jason Borenstein, *Society Affiliate, IEEE*,  
Justin Biddle, and Kelly Laas

**Abstract**—In recent years, numerous public, private, and non-governmental organizations (NGOs) have produced documents addressing the ethical implications of artificial intelligence (AI). These normative documents include principles, frameworks, and policy strategies that articulate the ethical concerns, priorities, and associated strategies of leading organizations and governments around the world. We examined 112 such documents from 25 countries that were produced between 2016 and the middle of 2019. While other studies identified some degree of consensus in such documents, our work highlights meaningful differences across public, private, and non-governmental organizations. We analyzed each document in terms of how many of 25 ethical topics were covered and the depth of discussion for those topics. As compared to documents from private entities, NGO and public sector documents reflect more ethical breadth in the number of topics covered, are more engaged with law and regulation, and are generated through processes that are more participatory. These findings may reveal differences in underlying beliefs about an organization's responsibilities, the relative importance of relying on experts versus including representatives from the public, and the tension between prosocial and economic goals.

**Index Terms**—Artificial intelligence, ethics, social implications of technology.

## I. INTRODUCTION

ARTIFICIAL intelligence (AI) is beginning to revolutionize numerous sectors of society, from research and transportation to finance and health care. Its near-term economic impacts are estimated to be in the trillions [1], and it is considered to be central to the Fourth Industrial Revolution

[2]. Its potential transformative impacts have led to a significant increase in attention to AI's social and ethical implications. As a result, over recent years, many organizations have produced documents that examine AI's ethical implications, articulate principles and guidance, and identify strategies to develop and implement AI responsibly. These documents – ethics codes, principles, frameworks, guidelines, and policy strategies – reflect the ethical viewpoints and priorities of leading organizations around the world. These include national governments, intergovernmental bodies, multinational corporations, prominent NGOs, and organizations created with a specific focus on AI.

Scholars have begun to analyze the content of these AI ethics documents. Some have used qualitative methods to identify themes across documents [3]–[7] or to support comparative analyses [8]–[10]; others have employed quantitative content analysis for similar reasons [11]. Still others have discussed second-order themes, such as the ethical assumptions underlying such documents [12] and the gap between ethical principles and actual practices [3], [7], [13]–[15]. Overall, the plurality of this work has focused on conceptually categorizing ethics topics and reducing them into a small number, typically 5–10, of core topics [6].

Jobin, Ienca, and Vayena (2019) have, for example, identified transparency, justice, fairness, nonmaleficence, responsibility, and privacy as concerns that typically appear in their set of 84 documents. Fjeld et al. (2020) identified eight similar principles in their analysis of 36 documents. Floridi and Cowls (2019) argued that the 47 AI ethics principles they reviewed fall within the traditional bioethics principles of beneficence, nonmaleficence, autonomy, and justice, along with a novel principle of explicability. In short, the primary thrust and focus of the prior literature has been to describe to what degree a global consensus around AI ethics is emerging.

Manuscript received May 27, 2020; revised October 9, 2020 and November 25, 2020; accepted January 3, 2021. This work was supported in part by the Science, Technology, and Innovation Policy Program, Georgia Institute of Technology.

© 2021 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Daniel Schiff is with the Georgia Institute of Technology, School of Public Policy, Atlanta, GA, U.S. (e-mail [schiff@gatech.edu](mailto:schiff@gatech.edu)).

Jason Borenstein is with the Georgia Institute of Technology, School of Public Policy and Office of Graduate Studies, Atlanta, GA, U.S. (e-mail [borenstein@gatech.edu](mailto:borenstein@gatech.edu)).

Justin Biddle is with the Georgia Institute of Technology, School of Public Policy, Atlanta, GA, U.S. (e-mail [justin.biddle@pubpolicy.gatech.edu](mailto:justin.biddle@pubpolicy.gatech.edu)).

Kelly Laas is with the Illinois Institute of Technology, Center for the Study of Ethics in the Professions, Chicago, IL, U.S. (e-mail [laas@iit.edu](mailto:laas@iit.edu)).

It is important to ask whether a global consensus in AI ethics is emerging; however, an emphasis on consensus can also mask important points of difference and omission. Differences in organizations' treatment of AI ethics can reflect competing priorities and translate into meaningful variation in actions. Even if organizations agree on the importance of a specific set of ethics topics, they might disagree about how particular ethical values should be interpreted, or how they should be prioritized relative to one another. They might also disagree on whether they are responsible for ensuring that a particular topic is addressed (e.g., privacy) or on processes of ethical oversight (e.g., governmental regulation versus self-regulation). Additionally, even if organizations agree to prioritize some set of ethics topics, all or most of these organizations might omit other ethics topics that deserve attention.

Our study is designed to address this gap in extant research by analyzing not only similarities, but also differences – in particular, differences across organizational sectors, conceived of here as the public sector (i.e., national governments or intergovernmental bodies like the European Union or United Nations), the private sector (mostly for-profit corporations), and the non-governmental sector (e.g., academic, advocacy, or professional organizations). To do so, our study looked at English-language documents published between January 2016 and July 2019 that include an examination of AI and ethics. We included document types such as ethics principles, frameworks, guidelines, and policy strategies, but not traditional academic research, opinion articles, or speeches. We focused on documents that examined AI generally; documents that targeted issues more broadly (e.g., data ethics) or narrowly (e.g., the ethics of autonomous vehicles) were excluded from our dataset.

After applying our inclusion/exclusion criteria, our collection included 112 AI ethics documents from public, private, and NGO organizations around the world. Fig. 1 depicts the final sample by country.

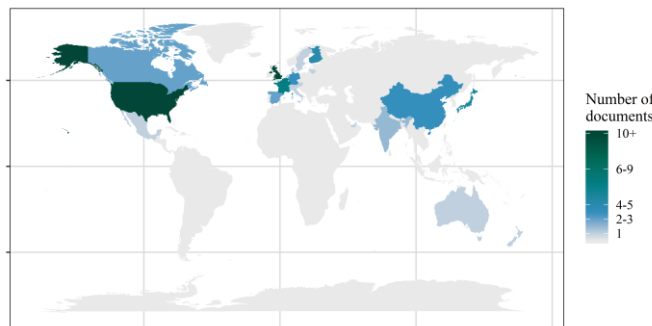


Fig. 1. AI ethics documents included in analysis by country.  $n=112$  documents. Country indicated is the international headquarters of the first authoring organization.

Previous studies typically identified 5-10 core ethics topics and coded each topic as either present or absent. In contrast, part of the uniqueness of our study and its associated analysis is that it includes a larger and more fine-grained set of ethics topics, 25 in total. Moreover, our coding system is not binary,

which allows for a more nuanced understanding of similarities and distinctions across different organizational sectors. Thus, while the literature has primarily looked for areas of topical consensus in AI ethics and has done so by distilling and condensing ethical concepts, we deliberately sought to determine whether divergences might be appearing as well.

Further, our study incorporates a sizable number of documents from the public sector, whereas much prior research in this field has concentrated on NGO and private sector ethics documents. Finally, our dataset is amongst the largest to date, with 112 documents. The identification of dissensus was therefore enabled by our use of a more fine-grained taxonomy of ethics and coding schema, by our robust sample of documents, and by our quantitative analyses that take advantage of these features.

In particular, we separately analyzed each document for 25 ethics topics from a taxonomy created by the authors. We also measured sectoral differences by coding for the degree of participatory engagement in the document's creation and level of engagement with law and regulation.<sup>1</sup> To improve coding reliability, a pair of researchers (the first author and another member of the research team) coded each document. After separately coding each document, the pair of coders held meetings to reconcile any differences and brought any unresolved disagreements to the full team.<sup>2</sup> The resultant dataset allowed us to review similarities and differences across organizational sectors quantitatively.

We have identified some striking sectoral differences. First, NGO and public sector documents have more participatory processes in the creation of documents, as well as more engagement with issues of law and regulation. Second, NGO and public sector documents reflect more ethical breadth and depth, according to measures we propose, and are generally more similar to each other than to private sector documents. Third, along with several areas of consensus, there are also conspicuous differences in the ethical priorities across sectors. While the private sector tends to emphasize ethical issues with ostensible technical fixes, such as algorithmic bias and transparency, the NGO sector addresses a wider set of topics such as accountability and misinformation, and the public sector focuses on unemployment and economic growth.

Several possible explanations may account for these differences in priorities surrounding ethics topics. Organizations might have different perceived scopes of responsibility given their traditional roles and audiences; they might view the role and relative importance of experts and the public in shaping decisions differently, and organizations may have varied motivations [26]. Our Discussion reviews these explanations further in terms of how 'ethical frames' may differ across sectors or organizations, contributing to dissensus in ethical priorities and translating into different courses of action.

Our study therefore seeks to answer the following question: *what similarities and differences can be detected in AI ethics documents produced by the public, private, and NGO sectors?* More generally, our primary research aim is to provide a fuller

<sup>1</sup> We assessed each document for additional variables, including publication date, location, and any discussion of ethics topics not covered in our taxonomy. We also reviewed each document for discussion of 17 policy sectors. Findings related to policy sectors are not included in this paper.

<sup>2</sup> More detail about our approach to document search and selection, the inductive identification of ethics topics, and our coding and analysis strategy is presented in the Methods section and in the Appendix.

understanding of how organizations from different sectors are addressing AI ethics in their published documents, including in terms of their ethical priorities, levels of engagement with law and regulation, and participatory involvement. We also provide a new public document-level database of AI ethics which can support a variety of analyses.<sup>3</sup> As organizations continue the work of translating ethical principles into practices, researchers would do well to focus on the organizational or sectoral contexts that are shaping AI ethics, and how these contexts might guide ethical priorities and actions.

## II. METHODS

### A. Document Collection

The data for our study is in the form of ethics codes, principles, frameworks, guidelines, policy strategies with ethics sections, reports, and other AI ethics documents. Our collection focused on documents drafted by public, NGO, and private sector organizations. Much of this gray literature is not typically found in academic databases [16].

We began the document collection process in Fall 2018 through web searches and citation scanning, and benefited from existing collections of AI ethics documents.<sup>4</sup> We engaged in ongoing monitoring of social media and news regarding AI ethics documents until early 2020. In early 2020, to ensure our dataset was robust, we examined the citations in two of the most comprehensive collections, a set of 84 documents from Jobin, Ienca, and Vayena (2019) and PricewaterhouseCoopers' collections listed in the 2019 AI Index Report [17]. Our initial sample consisted of 224 documents. We then applied our inclusion/exclusion criteria, outlined in Table 1.

TABLE I  
INCLUSION/EXCLUSION CRITERIA FOR SAMPLE IDENTIFICATION

| Criteria      | Inclusion  | Exclusion  |
|---------------|--|--|
| Document type | Codes, principles, frameworks, guidelines, policy strategies, reports                                | Academic research, opinion articles, drafts, speeches, standards, laws                     |
| Author        | Public sector, private sector, or non-governmental organizations                                     | Authors not acting as representatives of organizations                                     |
| Availability  | Publicly available   |  |
| Time frame    | January 2016 to July 2019  |  |
| Language      | English-language or high-quality translation   |  |
| Content focus | Focus on the intersection between ethics and AI, machine learning, autonomous or intelligent systems | Too broad: e.g., technology ethics or data ethics<br>Too narrow: e.g., autonomous vehicles |

First, we allowed for a variety of document types such as ethics codes, principles, frameworks, guidelines, policy

strategies, and reports. In most cases, traditional academic research, opinion articles, speeches, industry standards, or laws were excluded. The rationale for this approach was to include documents that were authoritative expressions of an organization's AI ethics principles.

Next, we decided to focus on three organization types: public, private, and NGO. We found that coding organization types at a finer level of granularity (e.g., trade unions or professional associations) was less reliable. The selected categories provided the sample sizes needed to support robust quantitative analyses and interpretation. Documents had to be publicly available and published between January 2016 and July 2019, when we halted the document collection process. Because of our own language limitations, we only included documents written in English or with an official or unofficial translation that we deemed of sufficient quality. If multiple versions of a document were available, we minimized redundancy by including only the latest within the time frame.

Finally, documents had to have a focus on both AI and ethics, which was assessed through document titles and text. Here, we defined AI operationally; if document authors perceived their work to be about AI (or used highly related concepts such as machine learning), we accepted it as such.<sup>5</sup> The document did not have to be exclusively about ethics, but it had to include significant discussion or sections devoted to ethics. As part of our screening process, we determined that some documents had too broad of a focus, such as data or technology ethics, or too narrow of a focus, such as the ethics of autonomous vehicles alone. These documents typically did not examine AI ethics directly, were less comparable to other documents, and so were excluded. Two coders independently assessed whether each document addressed both AI and ethics; disagreements were brought to the full team for discussion.

As outlined in Fig. 2, we identified 224 candidate documents in total, with a final sample of 112 after screening.

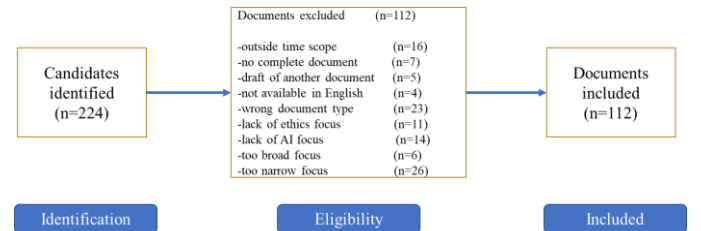


Fig. 2. Data collection and screening process.

### B. Development of the Codebook

We employed qualitative content analysis as our core methodology. The four-member research team began in August 2018 by reading a small number of AI ethics documents. Based on these documents and our own knowledge of AI ethics, we inductively developed an initial list of codes, including a preliminary set of ethics topics [18]. We then engaged in simple reliability testing, with all coders asynchronously coding two documents and comparing results [19]. We discussed challenges with the coding methodology and iteratively

<sup>3</sup> Database is available at <https://dx.doi.org/10.21227/fcdp-pa48>

<sup>4</sup> Document collections from Tim Dutton, the Future of Life Institute, and Nesta were especially helpful.

<sup>5</sup> While AI can be defined and understood in complex and inconsistent ways, we view this ambiguity as part of the context underlying the research.

improved our codebook and approach, which we then applied to a second set of two test documents. The process of refinement involved identifying codes that needed to be added, merged, separated, or defined more carefully. We also solicited feedback from a few external experts. Our final codebook as of the summer of 2019 includes variables for document name, authoring organizations, date of publication, organization sector, degree of public participation, engagement with law/regulation, 25 ethics topics, 17 policy sectors (not analyzed in this paper), additional open-ended data, and document-level variables used for data management purposes.

The variables for public participation and engagement with law and regulation are important for this analysis. The former variable was selected because an organization's willingness to seek diverse and public input may influence its determination of ethical priorities, as well as its beliefs about to whom it is accountable. Engagement with law and regulation is essential because it can serve as an (imperfect) proxy for how much an organization is attending to the practical implementation of ethical principles.

Other categories that might constitute variables are certainly worth considering; for example, Hagendorff (2019) examines the proportion of women amongst document authors and whether there is discussion of technical implementation. Indeed, in the development of our codebook, we considered coding additional categories, such as whether an organization proposed a monitoring or evaluation strategy, and if so, if it is internal or external to the organization. Yet we found some of these aspects to be largely absent or difficult to code reliably – arguably a finding in itself. While the variables we considered cannot tell a complete story about sectoral differences in AI ethics, we believe they are important nonetheless.

The choice of which ethics topics we included in our taxonomy is central to the study. How topics are defined and parsed is, to some degree, subjective and necessarily involves trade-offs. On the one hand, a taxonomy of ethics topics that is small potentially conflates distinct ideas. Our study, for example, better distinguishes between different types of harms that can result from AI, such as psychological harms from social media abuses and harms to democracy from misinformation, whereas other taxonomies collapse these into a single category. On the other hand, a list of ethics topics that aims to separate out every unique concept may generate unhelpful proliferation.

As two members of our team are academic philosophers, and one is a librarian with expertise in ethics codes, we undertook considerable efforts to identify conceptual distinctions and overlaps that presented a challenge for coding. For example, bias and inequality are related concepts: each can lead to the other. However, these topics were often contextualized and discussed separately within AI ethics documents, suggesting that many document authors conceived of them as distinct concepts. As a result, we found it essential to pay attention to how organizations were combining, separating, and implicitly or explicitly defining ethics topics. Our codebook thus reflects a balance between conceptual clarity and actual application.

We found that the 25 ethics topics in our final codebook provided robust coverage across many documents, led to

reliable coding, and addressed key issues not covered in the literature on AI ethics documents. Nevertheless, readers concerned with conceptual clarity and reliability should closely review our list of ethics topics to see how some topics might be merged together or disaggregated, potentially leading to new analyses.<sup>6</sup> We acknowledge that we may have missed important topics. Therefore, for each document, we captured an open-ended list of additional ethics topics not systematically captured by our 25 standard topics (see Table 2 in the Appendix).

### C. Coding Strategy

The coding process began in July 2019. The team entered all codes into a shared database; we did not use any qualitative analysis software as our approach was at the level of documents, though we did capture page numbers associated with coding choices. Below, we describe the coding criteria for some of our key variables.

We scored the 25 ethics topics as 0, 1, or 2. 0 refers to a topic that is absent. 1 refers to a 'minor' topic, mentioned in passing, left undeveloped, or otherwise not meeting the criteria for a substantive topic. 2 refers to a 'substantive' topic. A substantive topic has a whole section or principle devoted to it or is developed through discussion of two or more of the following: ethics topic definition, relevance/importance, and implementation/recommendations. Our coding strategy did not penalize documents, such as a single page ethics code, for their length. Therefore, a one-sentence ethics principle in a short document would be coded as a substantive topic if it was a clear expression of one of the 25 ethics topics. In other words, "substantive" was determined on a relative rather than absolute scale.

We coded the participation variable similarly, scoring each document as 1, 2, or 3. A document scored as 1 ('closed') was written by a single organization or limited body (e.g., a small private committee). A document scored as 2 ('semi-open') was authored by several organizations or a small to moderate number of people, possibly in a semi-private forum and with limited expert testimony. A document scored as 3 ('participatory') was composed or advised by multiple organizations and/or a large number of people, possibly through multiple workshops, conferences, a public comment period, and/or multiple iterations of the document.

Evidence for participation was drawn directly from the details within each document; to preserve consistency, we did not solicit additional information from organizational authors or other sources. This means it is possible that some documents were created with more participatory processes than we are aware of, which is a limitation of our approach. Nevertheless, how organizations choose to describe the creation process for their work is important, including whether they discuss the kinds of participation that were involved.

The engagement with law/regulation variable was coded as 0, 1, or 2. A document scored as 0 ('absent/generic') merely noted the general importance of regulations or failed to discuss the matter altogether. A document scored as 1 ('referenced') identified a particular law or regulation related to AI (such as by stating that the EU's General Data Protection Regulation

<sup>6</sup> We do not focus on how documents define ethics concepts in consistent or inconsistent ways, or what this might imply. Readers interested in the variety

of ways that documents treat ethics topics will enjoy Jobin (2019) and Fjeld et al. (2020). Future work is needed to explore the significance of such differences.

(GDPR) might be important to consider), but did not explore its application beyond generalities. A document scored as 2 (‘engaged’) included discussion of at least one law or regulation and its specific application to AI ethics issues.

In terms of organization sector, public sector organizations refers to government bodies and intergovernmental entities. Private sector organizations were primarily for-profit corporations. Non-governmental organizations include professional associations, research and advocacy groups, and various academic or other collaborations, such as those resulting from conferences or workshops. While some universities are technically public or private institutions, we found that these documents were more similar in kind (process, goals, and scope of responsibility) to NGO documents than either public or private documents and categorized them as such. In cases with multiple authoring organizations, we attributed a document to the first authoring organization. However, there are cases with authors from multiple sectors, or authoring organizations that do not easily fit into our public/private/NGO taxonomy.

#### D. Reliability

Ensuring the reliability of coding was a major concern, as documents engaged with ethics topics in a wide variety of ways. For example, one document might define justice in reference to the legal and criminal justice system, while another document focused on more abstract notions of fairness and equality. Another example is that some documents might discuss issues of ‘bias and fairness’ in a way that clearly incorporates concerns with long-term inequality, while other documents might focus on bias in algorithms without explicitly discussing downstream effects. The favored terminology also appeared to differ across cultural lines, regional contexts, and disciplinary influences. Therefore, we took several key steps to promote reliability in our coding strategy.

First, each document was coded separately by two team members, followed by a reconciliation process. In the reconciliation process, involving more than 40 hours of discussions, the two coders for each document discussed and attempted to resolve discrepancies. If disagreements persisted, or the two coders felt uncertain, questions were brought to the full research team. While this process increased the time and complexity of the analysis process, it was critical to increasing coding reliability.

Second, the first coder for all 112 documents was the same person – the first author of this paper. The second coder for a document rotated among the three additional members of the team. Having the same primary coder on all documents ensured that coding strategies did not begin to deviate across different pairs of researchers. Third, we took a low inference approach to coding, only attributing the text to a certain topic if relevant keywords were explicitly stated, or if we thought the discussion was unambiguous. For example, a statement that AI may harm some groups in society more than others could be safely attributed to the ‘inequality’ topic, but not to the ‘vulnerable populations’ topic, unless a specific vulnerable group per our definition was directly mentioned.

#### E. Analysis and Robustness Tests

We performed analysis and created most graphics using the R statistical suite (version 3.6) available in RStudio (version 1.2.5). This mostly involved calculating descriptive statistics such as mean scores for ethics topics by organization sector, and simple bivariate regressions reported in Fig. 7. This analytical strategy thus involved some quantitative content analysis. Of note, some documents could arguably be attributed to a sector other than the one we assigned to it. As a robustness check, we recalculated primary statistics in the main body of the paper: they do not vary meaningfully when a small number of organizations are moved to other sectors.

### III. RESULTS

Our results are striking, and they reveal potentially concerning trends in the landscape of AI ethics documents. They suggest that our fine-grained ethics taxonomy was a fruitful lens that uncovered patterns not readily visible when using less detailed methods or when focusing on consensus alone. We observe both stark and more subtle variation in the priorities of organizations beyond areas of consensus, including which ethical concerns certain sectors intentionally or inadvertently deemphasize or omit.

We first describe how documents vary by organization sector in terms of the level of participation for each document’s creation, and how deeply each document engages with questions of law and regulation. There are prominent differences here. Next, we describe areas of consensus in ethical focus across sectors. Afterwards, we report measures of breadth and depth of ethical coverage by sector, which also reveal stark differences. Next, a key part of our analysis is the identification of sectoral differences in the absolute and relative rankings of the 25 ethics topics. Finally, we note ethics topics that are often omitted before considering explanations for specific priorities and differences that we observed in the Discussion.

#### A. Participatory Engagement

Broad and public participation is significant in the sense that it may impact how an ethics discussion is framed. For example, a more diverse range of perspectives could lead to a wider and deeper consideration of AI ethics topics. On the other hand, closed processes could reflect an ‘expert’ orientation that “draw[s] a narrow circle of who can or should adjudicate ethical concerns around AI” [12].



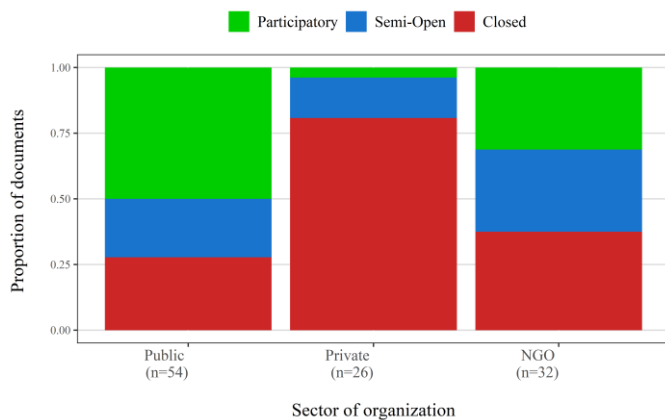


Fig. 3. Participatory engagement by organization sector. A relatively high percentage of public sector documents are created through participatory or semi-open processes, whereas most private documents have closed processes.

Fig. 3 depicts the results from all 112 documents, representing participatory engagement in terms of individuals or groups involved in a document’s creation. The results indicate that the NGO and public sector rely far more frequently on participatory processes compared to the private sector. Approximately three-quarters of public sector documents were produced through a semi-open or participatory format, with one-half through highly participatory processes. In contrast, broader participation seems to be lacking in the development of approximately three-quarters of private sector documents.

The public sector results are not wholly surprising, as public organizations often seek broad expertise through task forces, hearings, or other similar activities. Interestingly, the NGO results resemble those of the public sector: the NGO sector appears to have emphasized participatory processes as it organizes its thinking around AI ethics. This is potentially encouraging: as evidenced by a number of public sector documents [20], [21] that cite principles and frameworks developed by NGOs, ideas generated through participatory processes in one sector may permeate into another.

Yet increased participation does not guarantee that an AI ethics document is of higher quality, as poorly designed participatory practices and “participation-washing” may undermine benefits that increased participation brings [22]. Nonetheless, despite growing calls for diverse and public participation, including in many private sector documents studied here, e.g., [23], [24], the private sector does not seem to have realized these goals in the creation of their documents.

### B. Engagement with Law and Regulation

Like participatory processes, engagement with law and regulation can be an indicator that an organization has thought more deeply about translating AI ethics principles into practice. Organizations that discuss the law in depth often addressed more detailed changes to practice rather than emphasizing high-level principles and aspirations. As such, sectoral variation here could reflect differing sectoral motivations and strategies with important implications for organizations’ future actions. Yet a

lack of engagement with law could reveal a lack of seriousness about AI ethics, akin to what some have termed ‘ethics washing’ or ‘ethics shirking.’ [25]. For example, private organizations could be strategically advancing self-regulation in response to AI ethics as a way to avoid public oversight [26].

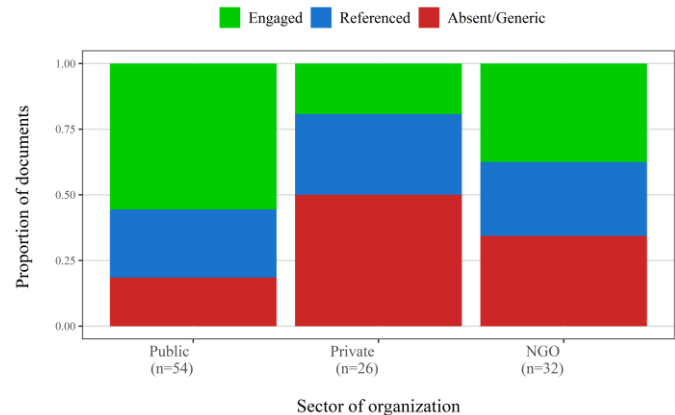


Fig. 4. Engagement with law by organization sector. A relatively high percentage of public sector documents are engaged with issues of law and regulation, compared to private sector documents.

In that light, it is potentially concerning that the results for engagement with law and regulation are similar, though less pronounced, as compared to results about participatory practice (Fig. 4). While more than half of public sector documents were ‘engaged’ in detail with law, only one in three NGO documents and one in five private sector documents moved beyond a cursory mention of a law or included no mention at all.<sup>7</sup>

However, engagement with law is an imperfect proxy for an organization’s level of seriousness about AI ethics. For example, results could merely indicate differences among perceived organizational scopes of responsibility. AI ethics documents from public sector organizations were often deliberately seeking to lay the groundwork for future regulations. In contrast, private sector organizations tended to be more focused on compliance with existing laws and informing external audiences, especially prospective customers, about their ethical priorities and internal business practices. However, even these assumptions about the scope of responsibility of organizations can be scrutinized, as we consider later on.

Moreover, organizations can convey seriousness about AI ethics in other ways, such as through engineering and computing practices, hiring or training policies, funding and project development decisions, or other organizational processes. Yet our codebook development process did not result in our inclusion of these proxies for operationalizing principles into practice, as these candidate topics were addressed inconsistently, vaguely, and rarely across documents and sectors. This absence, which may itself be a finding, led us to emphasize law and regulation.

In sum, findings reveal that private organizations are less engaged in participatory processes while developing their documents or in substantive discussions of law and regulation. While approaches to AI ethics are evolving, these initial results

<sup>7</sup> Notably, private sector documents were shorter (on average 14 pages) compared to NGO (46 pages) and public sector documents (39 pages). These differences could account for some of our findings here. Nevertheless, the

decision regarding how to *present* one’s AI ethics priorities and strategies is meaningful, as are decisions on whether to employ participatory processes and to engage with the law.

may nevertheless indicate key differences in the organizational contexts, motivations, and intended future actions.

### C. Consensus Topics

We now move to the key questions of ethical consensus and dissensus. Fig. 5 presents all 25 ethics topics from our team's taxonomy, averaged across organizational sectors, and presented in order from highest to lowest average score. The methodology section and Table II in the Appendix provide more detail on the identification of these 25 ethics topics.<sup>8</sup>

The five most prominent topics, according to our taxonomy, appearing in the document set are: social responsibility<sup>9</sup>, transparency, bias & fairness, privacy, and safety & reliability. Though our ethics topics are more finely parsed than in other studies, our results are nevertheless highly consistent with those studies, including Jobin, Ienca, and Vayena (2019), Fjeld *et al.* (2020), and Floridi and Cowls (2019). Since others have covered consensus topics in some depth, we do not do so here.

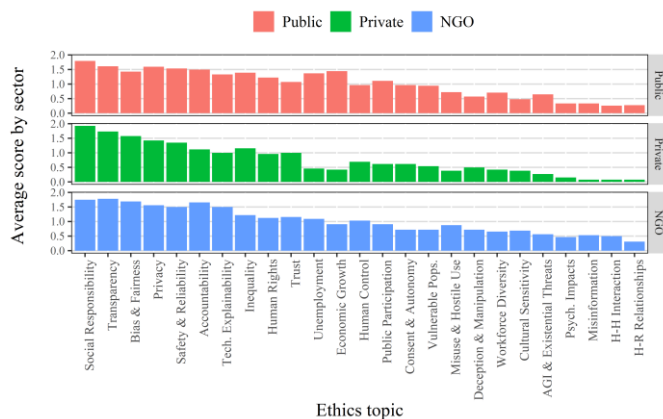


Fig. 5. Ranking of ethics topics by organization sector. Topics are presented in order from highest overall mean score. Ethics topics absent from a document were coded as 0; minor topics as 1; substantive topics as 2. Number of documents: Public (54); Private (26); NGO (32). Fig. 5 indicates several consensus topic across sectors as well as differences and omitted topics.

Our analysis, which may cover the largest number of AI ethics documents to-date, lends further credence to the notion that a global consensus is emerging on the importance of these five topics. Moreover, our results contribute to the literature by demonstrating this consensus quantitatively, and by showing that the consensus topics identified as important are viewed as such across public, private, and non-governmental sectors.

### D. Ethical Breadth and Depth

Another way of characterizing an organization's ethical lens, priorities, and seriousness is to evaluate its overall ethical 'breadth' and 'depth.' As Fig. 5 illustrates, the public sector and NGO sector have a wider breadth of ethical concerns, indicated by more consistently higher scores across the 25 ethics topics.

<sup>8</sup> AI ethics topics and concepts have been aggregated, divided, and parsed in various ways by researchers. No taxonomy of ethics topics can be fully comprehensive; any taxonomy involves trade-offs between different advantages and disadvantages for analysis. Understanding specific approaches to establishing a taxonomy is important for making sense of how ethics topics relate to one another, including differences, overlaps, and implications of any patterns.

In the private sector, fewer ethics topics are addressed on average, and scores drop off more rapidly.

We examine these trends more concretely by constructing quantitative measures of ethical breadth and depth, presented in Figs. 6 (a) and (b). Ethical breadth refers simply to the number of ethics topics from 0 to 25, including both minor and substantive topics. Ethical depth refers to the average score across all ethics topics, where a score closer to 2 indicates that more of the topics are considered to have a 'substantive' discussion in accordance with the coding scheme.

The ethical breadth measure in Fig. 6 (a) shows that the public sector addresses about one more ethics topic on average (16.6) than the NGO sector (15.5), while the NGO sector's breadth is similar to the global mean across all documents (15.2). Meanwhile, the private sector covers approximately 3 fewer topics on average (12), approximately one standard deviation lower than the global mean. This difference indicates that private organizations do have a more limited area of focus. This finding is also consistent with our expectations that a lower degree of public participation in the creation of the AI ethics documents could engender less robust consideration of AI ethics, though other explanations are possible. Indeed, simple bivariate regression results presented in Fig. 7 in the Appendix confirm that low participation and engagement with law are associated with less ethical breadth and depth.<sup>10</sup>

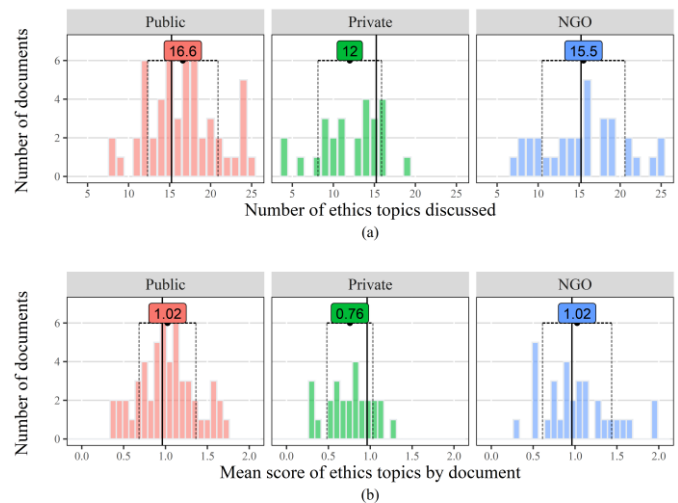


Fig. 6 (a). Breadth of ethical discussion by organization sector. Global mean (15.2) indicated by vertical line. Fig. 6 (b). Depth of ethical discussion by organization sector. Global mean (0.96) indicated by vertical line. For both subfigures, group means plus and minus one standard deviation are indicated with labelled box and dotted lines. These subfigures indicate higher levels of ethical breadth and depth in the public and NGO sectors.

In terms of ethical depth (Fig. 6 (b)), the public and NGO sectors (both 1.02) have a similar depth to the global mean (.96), while the private sector has less depth (.76), again about one standard deviation lower than the global mean. Also notable is

<sup>9</sup> Social responsibility is similar to what Floridi and Cowls refer to as 'beneficence' and could also be understood as related to public good, welfare, or well-being.

<sup>10</sup> See Appendix IV for a discussion of why these correlational findings are inconclusive. More work is needed to understand the causal pathways here.

that most private sector documents are concentrated in the .5 to 1.0 range, while there is visibly greater variance in the other sectors.<sup>11</sup> The regression results presented in Fig. 7 provide further correlational evidence of these sizable sectoral differences: the private sector, but not NGO sector, has a significantly lower breadth and depth of ethical discussion.

Overall, these results suggest that private sector organizations have a narrower lens of ethical focus as well as less depth in terms of ethical discussion, which may convey more superficial engagement with AI ethics issues. Results also indicate that the profile of NGO organizations appears to be broadly similar to the public sector along these measures. While these high-level measures can convey meaningful information and may be useful measures for future AI ethics research, there are additional patterns worth disaggregating which we discuss next.

### E. Prominent Sectoral Differences

A key focus of this empirical work is to identify differences in ethical priorities that may be masked by other conceptual and analytical approaches. To assess differences across organizational sectors, we calculated differences in relative rankings of ethics topics (ranging from 1 to 25) as well as absolute differences in mean scores. Based on these patterns, we identify a few key findings along with possible interpretations of the data.

First, the public sector has a visibly higher emphasis on issues of unemployment and economic growth than does the NGO sector and especially the private sector. Such an overriding concern with macro-economic issues may suggest that an economic logic, often attributed to firms, may importantly characterize public sector motivations as well. These results could also be explained by differences in perceived scope of responsibility: public entities such as governments may view their responsibility as pertaining to the broader public while private companies might focus on their shareholders and clients. Yet the presumed appropriate scope of responsibility for each organization is open to normative scrutiny; for example, one could argue that the public sector is overly focused on economic growth, while the private sector does not sufficiently account for its role in possible unemployment related to AI.

Second, the NGO sector more consistently raises a broader range of ethics topics than both the public and private sectors. For example, in addition to the topics standardly discussed by each sector, the NGO sector is more focused on accountability, misuse and hostile use, misinformation, and explainability than the private sector. This is consistent with a ‘watchdog’ role for the NGO sector, where NGOs urge action by private or public sector actors. Further, compared to the public sector, which is especially attentive to ‘practical’ issues like unemployment, the NGO sector addresses topics that are perhaps more ‘philosophical’ like human-human interaction, bias & fairness, deception & manipulation, and cultural sensitivity.

Third, though the private sector is not typically higher than other sectors in absolute terms, it gives higher relative priority to social responsibility and trust, which suggests that the private sector wants to signal that it is taking these issues seriously [26]–[28]. The private sector also emphasizes issues such as bias & fairness, consent & autonomy, transparency, deception & manipulation, and vulnerable populations to a lesser extent. This may again result from the private sector’s perceived scope of responsibility that emphasizes clients and customers. A number of these topics could also be understood as the sorts of issues companies responsible for a bulk of AI development can purport to remedy with technical fixes. However, such a focus places “design as [the] locus of ethical scrutiny” [12], potentially at the expense of questioning business values or practices that should be within the bounds of conversation.

### F. Omitted Topics

An emphasis on consensus alone can draw attention away from omitted or neglected topics that may be worthy of ethical concern. Our analysis indeed reveals some ethics topics that are *not* prominent in global AI ethics discourse. The ethics topics from our taxonomy that received the least attention are cultural sensitivity, artificial general intelligence (AGI) & existential threats, psychological impacts, misinformation, human-human interaction, and human-robot relationships.

Two key questions about this finding are why certain topics are deprioritized and whether such choices are prudent. We reflect on but cannot definitively demonstrate why such choices were made. First, AGI & existential threats may be underemphasized because they are arguably more future-oriented, and because documents emphasized relatively near-term ethical concerns. Even those documents that *did* discuss AGI & existential threats largely argued that these ethical concerns were uncertain or unrealistic and, therefore, should not be viewed as a current priority [29], [30]. For those concerned with long-term safety and alignment of AI systems with human values [31], [32], this explicit deprioritization may be a source of serious concern.

Another commonly omitted set of topics were those that addressed social and relational issues. These included psychological impacts, impacts on human-human interaction, human-robot relationships<sup>12</sup>, and cultural sensitivity. These might be perceived as more indirect or subtle compared to prominent AI ethics issues like privacy or transparency, and may not yet have garnered the attention of the authoring organizations. These omissions could also be explained by the ethical frameworks and underlying assumptions used by organizations [33], such as ethical theories that emphasize consequences or duties rather than personal relationships or character traits.

While foci of attention change over time, including in AI ethics [34], the computing community should note that potentially critical social and ethical implications of AI, only

<sup>11</sup> However, this apparent lack of depth is partially explained by the smaller number of topics. When we calculate the depth of ethical discussion alternatively by excluding ethics topics indicated as absent, the average scores for all three sectors are more similar, around 1.5 to 1.63. For topics that *are* discussed then, ethical depth appears relatively similar.

<sup>12</sup> Human-robot relationships were defined by our taxonomy to refer narrowly to social-emotional relationships between humans and robots. In conjunction with that, documents focusing primarily on robot ethics were excluded from the document collection, which may also help to explain why the appearance of this topic was relatively rare.



some of which we identify here<sup>13</sup>, may not be fully gaining the attention of leading multinational organizations and governments. If these topics do not reach the agendas of policymakers and leading organizations, they may be undervalued in public debate, organizational change, and policy adoption. More work is needed to reflect on whether the priorities and omissions identified here are justified.

#### IV. DISCUSSION

To help make sense of our findings, we draw on the notion of ‘ethical frames,’ or a set of grounding assumptions about how ethical problems and solutions are defined and who is responsible for making these decisions. Our understanding of ‘ethical frames’ builds on that of Greene *et al.* (2019). We consider here how ethical frames may vary across sectors or organizations, as ethical frames may be shaped by technical and social arrangements, underlying values, and norms that govern how things are typically done in different contexts.

For example, organizations can differ in the processes they set for defining problems and solutions, such as by placing boundaries around who makes these decisions. These differences in processes can, in turn, lead to differences in prioritization, interpretation, and presentation of ethical issues. Organizations may knowingly or unknowingly ignore certain topics, which can have significant consequences. If the public sector pays less attention to cultural sensitivity, for example, conversations about the topic might not fully take root. Organizations can also define topics differently, such as by defining explainability in a way that places more stringent demands on AI developers [35], potentially opening organizations up to increased scrutiny and legal liability. What an organization says in an AI ethics document can therefore provide insight into how it frames ethical issues.

Below, we discuss responses to three overarching questions, which might help to reveal how organizations frame and approach AI ethics. Whereas Greene and colleagues describe several ethical frames that they believe are *shared* across the AI ethics documents they studied [12], we seek to draw attention to both similarities and differences that might appear across sectors. Our results suggest that there are important differences in focus and emphasis across the three sectors we examined.

##### A. Perceived Scope of Responsibility

The first question is: *What is the scope of issues that an organization is responsible for?* The answer to this question reveals boundaries around how an organization defines ethical problems and solutions, and also embeds the question of *to whom* an organization is responsible. While a focus on consensus in AI ethics documents may suggest a unified notion of collective responsibility, we think that there are visible differences in how they frame their scope of responsibility, which result from longstanding assumptions about organizations’ traditional roles.

For example, public sector documents focus on economic growth at the national level, while private sector documents

emphasize technically ‘fixable’ and customer-focused issues, and NGO documents emphasize private sector accountability and a broader range of ethical issues. In the case of the public sector, this is likely tied to the nature of its work. As representatives of government, these organizations may perceive responsibility to retain or establish a competitive edge at the national level. In the private sector, AI businesses may emphasize a fiduciary duty towards shareholders, or may be reacting to specific ethical concerns voiced about their products. NGO stakeholders may view their role as a ‘watchdog,’ in attempting to strategically influence private and public actors as an advocate rather than as a direct actor.

These differences raise key questions. First, are organizations ‘correct’ about the limits of their scope of responsibility? Along related lines, will the distribution of their perceived responsibilities allow for adequate coverage and response to the many ethical issues raised? For example, some private organizations express concern in their documents about broader societal issues like unemployment. Yet they may consider such issues as outside the scope of their purview and delegate the responsibility to solve these issues to public institutions. This conception of scope of responsibility helps to reinforce a focus on “design as [the] ethical locus of scrutiny” [12], limiting the responsibility of the private sector to “better building” of AI. Such a focus serves to ‘frame out’ other possible solutions. For instance, it may as a byproduct entail that neither business practices nor the broader socio-economic system are sufficiently examined. Yet is this the most effective way to address AI ethics?

Should the distribution of responsibility be reimaged to better address these issues? For example, it may be the case that businesses focus on innovation, and government is expected to address any unemployment or inequality that results. However, it could also be the case that businesses, individually and collectively, should think more proactively about mitigating unemployment, while the government acts to incentivize and encourage such prosocial behavior. In short, an organization’s perceived scope of responsibility constitutes a key part of its overall identity and ethical frame. Yet, given how AI is reshaping human life at local, national, and global scales, it may be time to reexamine traditional assumptions about an organization’s sphere of influence and scope of responsibility.

##### B. Perceived Role of Stakeholders

The next guiding question is: *Who should decide?* Responses to this question help constitute the frame around who is included – and excluded – in the process of defining ethical problems and determining the solutions to them. This frame, therefore, establishes who is perceived as a subject versus object of AI ethics problems and solutions. The question of *who should decide* raises the further question: do differences in participation shape ethical priorities? A broader sense of whose input counts could lead to a subsequently broader range of ethical concern, or one that more carefully responds to issues faced by the public or other stakeholders.

<sup>13</sup> We recognize that our coding strategy has missed ethical issues that are potentially important, such as responsible research practice. Table III provides detail on a wider set of ethical issues identified in our document collection but

not included in our taxonomy. These topics are often (rightly or wrongly) omitted in AI ethics discussion.

In response to the former question, Greene *et al.* argue that AI ethics documents present AI ethics as “a project of expert oversight” which serves to “draw a narrow circle of who can or should adjudicate ethical concerns around AI.” Meanwhile, the public is often viewed as “a body to be educated and surveyed” by the experts. Even when a broader range of stakeholders is brought in to advise on AI ethics, the stakeholders are often intentionally “limited in their capacity to impact design” [12]. Under this frame, participation serves as a mere legitimization exercise, meant to alleviate concerns and constituting a form of participation-washing [22].

Yet, we find that this question about inclusion in the decision-making process is answered differently across sectors. Public sector and NGO documents are predominantly driven by inclusive and participatory processes, whereas private sector documents are largely not. Do these differences in participation lead to divergences in ethical priorities? Our results suggest that they do, given the distinctly narrower ethical scope of the private sector and correlations between participation and ethical breadth and depth (see Appendix Fig. 7). Private sector documents attend to a smaller number of ethical topics, on average, and focus on a narrower scope of action.

To better understand the significance of this, however, it is critical to assess the kinds of participatory processes that are used across all sectors. Which are genuinely transformative as opposed to merely performative or otherwise ineffective? To what extent do diverse stakeholders have the influence to shape decisions, influence an organization’s thinking, challenge conclusions, and engage over the long term? Beyond examining which stakeholders are included in document creation as we do here, assessing the impact of participation or the lack thereof requires a fuller understanding of best practices for crafting meaningful participatory processes [38].

### C. Perceived Importance of Underlying Values

The last guiding question we consider addresses underlying motivations and asks: *Which values should drive an organization’s decisions?* A sizable amount of prior work views socially responsible or beneficent AI as a consensus value [4], [6]. Yet other scholars and commentators have specifically criticized the private sector and argued that its stated motivations are better understood as signaling strategies than as genuine signs of social responsibility [27], [28]. While assessing the true motivations of organizations, or the individuals and coalitions that constitute them, is difficult<sup>14</sup>, our results provide some evidence that there may be a gap between expressions of interest in social responsibility and a commitment to action.

Social responsibility is the highest-ranked topic across all sectors. Yet, in the private sector, where social responsibility and trust are disproportionately prioritized compared to other issues, we see lower levels of ethical depth and breadth, as well as less engagement with issues of law and regulation, one proxy for taking ethics seriously. Though our research is subject to several limitations, the divide observed in this study may be

explained by a signaling strategy that in part substitutes for genuine concern for social responsibility.

However, while the tension between economic and social values in the private sector has a long history [39], [40], our results also challenge single-minded attention to the private sector’s motivations. We find that the public sector greatly prioritizes issues of economic growth and unemployment as compared to the other sectors, and states this focus openly. Meanwhile, a corresponding increase in attention to issues such as inequality or human rights is lacking in the public sector. From these data, and our comparative review of the entire document collection, it is evident that the public sector is substantially driven by an economic logic, and usually more explicitly so than the private sector. Indeed, the primary stated goals of many public sector organizations surrounding AI emphasize innovation and economic growth, often aimed at promoting a nation’s economic, political, or even military status against competing nations [41], [42].

In contrast, ethical issues are often viewed as side considerations and, at times, even as barriers that might prevent successful or ‘trustworthy’ innovation. For example, public sector organizations often respond to potential labor displacement induced by automation by nevertheless promoting creative destruction as the favored solution [21], [43]. Consequences such as increased inequality are treated as unfortunate implications, only partially addressed by supply-side labor reforms like increased education and worker re-skilling. Meanwhile, alternative solutions like more radical reforms to social safety nets or rethinking around the future of work are given less attention. This framing of appropriate solutions is aimed at furthering innovation as the primary goal. We think these findings should encourage a shift in focus from inspecting the economic motivations of corporations alone to scrutinizing the broader economic-driven logic underpinning AI innovation, policy, and ethics altogether.

## V. LIMITATIONS

Importantly, documents included in our analysis must have had an English-language version, which entails that regions of the globe may be underrepresented. Though 25 countries were included in the dataset, the majority of these are high-income countries and other world powers. The reliability of our coding and analysis is also constrained by the sheer variation and complexity of the documents, including differences in regional, cultural, and stylistic orientation. The documents presented a wide variety of terms and surrounding discussion, often without explicit definitions, including terms which required interpretation to apply to our taxonomy.

Documents also referred to the same term in a variety of different ways. It is possible that the authors of documents did not have clear or unambiguous definitions in some cases. For example, our taxonomy included topics that were closely related, such as ‘bias’ and ‘inequality.’ At times, it was challenging to parse whether a document was addressing one or both concepts. Our efforts to improve reliability can only

<sup>14</sup> In prior work, we discuss six possible overlapping motivations that organizations have for creating AI ethics documents, which could explain sector variations in the treatment of ethics topics [26]. It is difficult to assign a

definitive motivation to an individual organization, though some trends at the sector level can be examined.

account for some of the complexity and ambiguity in these qualitative data. Moreover, we made a few refinements to our codebook while document coding was already underway. This included removing, merging, and renaming several ethics topics. Additionally, other important topics, such as research ethics, should have potentially been included in our taxonomy. Table III in the Appendix provides additional insight about common ethics topics that other researchers may wish to consider.

Our evidence of how organizations view AI ethics reflects only a subset of possible sources. Organizations may have other venues and documents that reveal their ethical thinking and behavior, such as internal working groups, organizational policies, software tools, and unpublished team meeting notes. The documents we coded do not reflect the entirety of any organization's ethical motivations, attitudes, or behavior. Indeed, many of the documents were works in progress, and organizations may have since made changes to their principles, practices, or leadership that our analysis did not capture.

Also, our choice to analyze participation and engagement with law does not present a complete picture, as these categories are among many others of possible importance. Yet as part of our codebook development and conceptualization, we did consider other categories, such as whether documents proposed monitoring or evaluation strategies, and whether any monitors were internal or external to the organization.<sup>15</sup> We overwhelmingly found that these discussions were absent, which may itself be a finding. Nevertheless, to build a more robust picture, researchers should assess additional ethical problems, solutions, and processes related to AI that may have importance to public, NGO, and private organizations.

## VI. CONCLUSION AND FUTURE RESEARCH

This study evaluated 112 AI ethics documents published between January 2016 and July 2019 from the public, private, and NGO sectors across 25 different ethics topics. Findings reveal clear differences in the treatment of AI ethics across organization types. We find that public and NGO documents are more participatory in their creation and more engaged with the law. Private sector documents appear to be more concerned with client and customer-related ethical issues that may lend themselves to a technical fix. Public sector documents emphasize economic growth and unemployment, and NGO documents cover a range of ethical issues that are given less attention by other sectors.

Yet our study leaves many unanswered questions that are worth examining.<sup>16</sup> These include how these documents define AI and particular ethical concepts, what similarities and differences are embedded in those definitions, and how this variation may shape ethical priorities and actions. This line of research should pay attention to divergences and omissions as well as areas of consensus. For example, do private sector organizations tend to define ethical concepts such as privacy or bias differently from NGOs? Another opportunity for future

researchers is to study categories discussed in AI ethics documents beyond participation and engagement with law and regulation, especially categories that address proposed solutions to AI ethics problems. Some areas of interest include funding priorities, hiring or training policies, computing and engineering design processes, and monitoring and assessment strategies. Studies that leverage multiple sources and methods could help answer how documents have or have not contributed to changes in industry practices and in policy adoption. For example, interviews of AI ethics document authors could provide more detailed insights into underlying motivations and goals, while impact assessments of AI systems could indicate whether ethical principles are being realized in practice.

While there is much still to understand, our findings help to reveal organizational differences in the ethical framing of several key issues, including about the perceived scope of responsibility, the role of experts and the public, and the tension between economic and other values. Though the current literature has revealed areas of consensus across organizations, our results indicate that sectoral differences are also worthy of attention, along with regional and cultural differences. Understanding these variations and unpacking assumptions can help scholars and stakeholders make sense of the distribution of responsibilities, identify opportunities for better collaboration, better guide efforts to improve internal practices and formal regulation, and expand the space of possible solutions.

## REFERENCES

- [1] J. Bughin, J. Seong, J. Manyika, M. Chui, and R. Joshi, "Modeling the global economic impact of AI," McKinsey Global Institute, Sep. 2018.
- [2] K. Schwab, *The fourth industrial revolution*, First U.S. edition. New York: Crown Business, 2016.
- [3] A. Jobin, M. Ienca, and E. Vayena, "The global landscape of AI ethics guidelines," *Nat. Mach. Intell.*, vol. 1, no. 9, pp. 389–399, Sep. 2019.
- [4] J. Fjeld, N. Achten, H. Hilligoss, A. Nagy, and M. Srikumar, "Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI," Berkman Klein Center for Internet & Society, Rochester, NY.
- [5] M. Gibert, C. Mondin, and G. Chicoisne, "Montréal Declaration of Responsible AI: 2018, Overview of International Recommendations for AI Ethics," University of Montréal, Dec. 2018.
- [6] L. Floridi and J. Cowl, "A Unified Framework of Five Principles for AI in Society," *Harv. Data Sci. Rev.*, Jun. 2019.
- [7] T. Hagendorff, "The Ethics of AI Ethics: An Evaluation of Guidelines," *Minds & Machines*, Feb. 2020.
- [8] T. Dutton, B. Barron, and G. Boskovic, "Building an AI World: Report on National and Regional AI Strategies," CIFAR, Toronto, Canada, Dec. 2018.

<sup>15</sup> Hagendorff (2020) similarly finds no discussion of technical implementation strategies in the overwhelming majority of documents he reviewed.

<sup>16</sup> We thank anonymous reviewers for suggesting some lines of inquiry.

- [9] A. Daly et al., “Artificial Intelligence, Governance and Ethics: Global Perspectives,” SSRN Scholarly Paper ID 3414805, Jul. 2019.
- [10] C. Cath, S. Wachter, B. Mittelstadt, M. Taddeo, and L. Floridi, “Artificial Intelligence and the ‘Good Society’: The US, EU, and UK Approach,” SSRN Scholarly Paper ID 2906249, Dec. 2016.
- [11] Y. Zeng, E. Lu, and C. Huangfu, “Linking Artificial Intelligence Principles,” ArXiv181204814 Cs, Dec. 2018.
- [12] D. Greene, A. L. Hoffmann, and L. Stark, “Better, Nicer, Clearer, Fairer: A Critical Assessment of the Movement for Ethical Artificial Intelligence and Machine Learning,” *Crit. Ethical Stud. Digit. Soc. Media*, p. 10, 2019.
- [13] J. Morley, L. Floridi, L. Kinsey, and A. Elhalal, “From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices,” *Sci. Eng. Ethics*, Dec. 2019.
- [14] D. Schiff, B. Rakova, A. Ayesh, A. Fanti, and M. Lennon, “Principles to Practices for Responsible AI: Closing the Gap,” arXiv:2006.04707 [cs], Jun. 2020.
- [15] B. Mittelstadt, “Principles alone cannot guarantee ethical AI,” *Nat. Mach. Intell.*, vol. 1, no. 11, Art. no. 11, Nov. 2019.
- [16] H. R. Rothstein and S. Hopewell, “Grey literature,” in *The handbook of research synthesis and meta-analysis*, 2nd ed, New York, NY, US: Russell Sage Foundation, 2009.
- [17] R. Perrault et al., “The AI Index 2019 Annual Report,” Human-Centered AI Institute, Stanford University, Palo Alto, California, U.S., Dec. 2020.
- [18] D. R. Thomas, “A General Inductive Approach for Analyzing Qualitative Evaluation Data,” *Am. J. Eval.*, vol. 27, no. 2, pp. 237–246, Jun. 2006.
- [19] K. A. Hallgren, “Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial,” *Tutor. Quant. Methods Psychol.*, vol. 8, no. 1, pp. 23–34, 2012.
- [20] Ministry of Science, Innovation and Universities, Spain, “Spanish RDI Strategy in Artificial Intelligence,” Mar. 2019.
- [21] Office of the Prime Minister, Malta, “Malta: Towards An AI Strategy,” Valletta, Malta, Mar. 2019.
- [22] M. Sloane, E. Moss, O. Awomolo, and L. Forlano, “Participation is not a Design Fix for Machine Learning,” arXiv:2007.02423 [cs], Aug. 2020.
- [23] KPMG, “Guardians of trust: Who is responsible for trusted analytics in the digital age?,” KPMG, Amstelveen, Netherlands, Feb. 2018.
- [24] Futurice, “The Futurice Principles for Ethical AI,” Futurice, Helsinki, Finland, Nov. 2018.
- [25] K. S. Gill, “Prediction paradigm: the human price of instrumentalism,” *AI Soc.*, vol. 35, no. 3, pp. 509–517, Sep. 2020.
- [26] D. Schiff, J. Biddle, J. Borenstein, and K. Laas, “What’s Next for AI Ethics, Policy, and Governance? A Global Overview,” in *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, New York, NY, USA, Feb. 2020, pp. 153–158.
- [27] E. Bietti, “From ethics washing to ethics bashing: a view on tech ethics from within moral philosophy,” in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, Barcelona, Spain, Jan. 2020, pp. 210–219.
- [28] R. Abebe, S. Barocas, J. Kleinberg, K. Levy, M. Raghavan, and D. G. Robinson, “Roles for computing in social change,” in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, Barcelona, Spain, Jan. 2020, pp. 252–260.
- [29] Executive Office of the President, National Science and Technology Council (NSTC), Committee on Technology, United States, “United States: Preparing for the Future of Artificial Intelligence,” Washington, D.C, Oct. 2016.
- [30] Austrian Council on Robotics and Artificial Intelligence, Ministry for Climate Action, Environment, Energy, Mobility, Innovation and Technology, “Shaping the Future of Austria with Robotics and Artificial Intelligence,” Apr. 2018.
- [31] C. Prunkl and J. Whittlestone, “Beyond Near- and Long-Term: Towards a Clearer Account of Research Priorities in AI Ethics and Society,” arXiv:2001.04335 [cs], Jan. 2020.
- [32] S. Russell, D. Dewey, and M. Tegmark, “Research Priorities for Robust and Beneficial Artificial Intelligence,” *AI Mag.*, vol. 36, no. 4, Art. no. 4, Dec. 2015.
- [33] T. Hagendorff, “The Ethics of AI Ethics: An Evaluation of Guidelines,” *Minds & Machines*, Feb. 2020.
- [34] Department of Homeland Security, Office of Cyber and Infrastructure Analysis, United States, “United States: Artificial Intelligence Risk to Critical Infrastructure,” Washington, D.C, Jul. 2017.
- [35] Y. Zhou and D. Danks, “Different ‘Intelligibility’ for Different Folks,” in *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, New York, NY, USA, Feb. 2020, pp. 194–199.
- [36] R. A. Ankeny and S. Leonelli, “Repertoires: A post-Kuhnian perspective on scientific change and collaborative research,” *Stud. Hist. Philos. Sci. Part A*, vol. 60, pp. 18–28, Dec. 2016.
- [37] A. Lakoff and S. J. Collier, “On regimes of living,” *Glob. Assem. Technol. Polit. Ethics Anthropol. Probl.* Malden Oxf. Carlton Blackwell Publ., pp. 22–39, 2005.
- [38] J. Glicken, “Getting stakeholder participation ‘right’: a discussion of participatory processes and possible pitfalls,” *Environ. Sci. Policy*, vol. 3, no. 6, pp. 305–310, Dec. 2000.
- [39] V. L. Nielsen and C. Parker, “Mixed Motives: Economic, Social, and Normative Motivations in Business Compliance,” *Law Policy*, vol. 34, no. 4, pp. 428–462, 2012.
- [40] L. J. Spence and R. Rutherford, “Social responsibility, profit maximisation and the small firm owner-manager,” *J. Small Bus. Enterp. Dev.*, vol. 8, no. 2, pp. 126–139, Jan. 2001.
- [41] State Council of China, “China’s New Generation of Artificial Intelligence Development Plan,” Beijing, Jul. 2017.
- [42] NITI Aayog, India, “National Strategy for Artificial Intelligence #AIFORALL,” NITI Aayog, Delhi, India, Jun. 2018.

[43] Ministry of Economic Affairs and Employment,  
“Finland’s Age of Artificial Intelligence Turning Finland into  
a leading country in the application of artificial intelligence,”  
Ministry of Economic Affairs and Employment, Helsinki,  
Finland, 47/2017, Dec. 2017.